



[illegible]



```
/codon_start=1
/product="tumor endothelial marker 8 precursor"
/protein_id="AAK52094.1"
/db_xref="GI:14017381"
/translation="MATAERALGIGFQWLSLATLVLCAGGGRRDGGPACVGGFD
LYFLDQSGVLHWHNEIYYEVLQAHKFIYSPQLRMSFVSTRTGLMKLTEDREQI
RQGLLEQLVPGDGTVMHEGFERASEQIYYENRQGYRTASVIALITDGEHDLFFY
SERANRSKDLGAVYCVGVKDFNETOLARADSKDHFVPVNDFOALQGIHSLTKK
SCIETLAEPSTICAGESFOVVVRGNGFRHARNVDRLVLCSEKINDSVTLNEKPSVED
TYLLCPAPILKEVGHKAALQVSMNDGLSFISSSVLIITTHCSGDSILAILLFLLL
ALALMFVPLCCITVILEVPPPAESEEEDDDGLPKKWPVTVDASYGGRGVGGIK
RMEYRWEKSGTEGAKLEKARNVRKMEQEFEPENLNMMNRSSPRKWSPI
KGKLDAWLKRGYDRVSNRPGDGTGRCINFTRVKNQNPAPYPLNNVHTSSPPP
APIVTPPPAPHPCHPPPSAPTPIPSPLPPPPQAPPNRPAPPPPPPPPSV"
144..224
sig_peptide
/gene="TEM8"
mat_peptide
/gene="TEM8"
225..1835
/misc_feature
/product="tumor endothelial marker 8"
1104..1172
/gene="TEM8"
/region="Region: predicted transmembrane domain"
1601 a 1306 c 1187 g 1446 t
BASE COUNT 1601 a 1306 c 1187 g 1446 t
ORIGIN
Query Match 98.8%; Score 1090.8; DB 9; Length 5540;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1092; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 atggccacggcgagcgagagccctgcgcagtcgctccagtcgctctctttggccact 60
DB 144 ATGGCCACGGCGAGCGGAGAGCCCTCGGCATCGCTTCCAGTGCTCTTTGGCCACT 203
QY 61 ctgggtgctcatctgcgcggcggaaggacgcaggagagatgggggtccagcctgctac 120
DB 204 CTGGTGCTCATCTGCGCGCGGGAAGGGGACGACGAGGAGTGGGGTCCAGCCTGCTAC 263
QY 121 ggcgattgacctgactctatttgacaaatcagagaagtgtactgcaccacaggaat 180
DB 264 GCGGATTGACCTGACTTTCATTGGAACAATCAGGAAGTGTCTGCACCACCTGGAAT 323
QY 181 gaaatctattcttgtgaaacagttggctcacaattcatcagccacagttgagaatg 240
DB 324 GAAATCTATTACTTGTGGAACAGTGTGGTCCACAAATTCATCAGCCACAGTTGAAATG 383
QY 241 tctttattgtttctccaccgcggagaaacacctaatgaaactgacagaagacagagaa 300
DB 384 TCCTTTATTGTTTCTCCACCGGAGAACACACTTAATGAAACTGACAGAAGACAGAAA 443
QY 301 caaatccgtcaaggcctagaagaactccagaagttctgcaggagagacacttacatg 360
DB 444 CAATCCGTCAAGGCTAGAAACTCCAGAAAGTCTGCCAGGAGGAGACACTTACATG 503
QY 361 catgaaggattgaaaggccagtgagcagatttattatgaaacacagaaaggttacagg 420
DB 504 CATGAAGGATTGAAAGGCCAGTGAGCAGATTATTATGAAACACAGACAAGGTACAGG 563
QY 421 acagccagcgcatcatctgtcttgactgaggaactccatgaagatctctttctat 480
DB 564 ACAGCCAGCGTCATCATCTTCTTGTACTGTAGTGTGAGAACTCCATGAAGATCTCTTTTCTAT 623
QY 481 tcagagagggagcctaataaggctcagagatctgtgcaattgtttactgttggtg 540
DB 624 TCAGAGAGGAGGCTAATAGTCTCGAGATCTTGGTGCATTTGTTTACTGTGTGTGTG 683
QY 541 aaagatttcaatgagacacagctggcccgattggggacagtaagatcatgtgtttccc 600
DB 684 AAAGATTTCATGAGACACAGCTGGCCGANTTGGCGACAGTAAGGATCATGTGTTTCC 743
QY 601 gtgaatgacggctttcagggctctgaaggcatcaccatcgaattttgaagaagtcctgc 660
DB 744 GTGAATGACGGCTTTTCAGGCTCTGCAAGGCATCATCCACTCAATTTTGAAGAAGTCTCTGC 803
```

```
QY 661 atcgaattttagcagctgaaccatccaccatattgtgcaggagagtcatttcaagttg 720
DB 804 ATCGAAATTCAGCAGCTGAACCATCCACCATTATGTGCAGGAGAGTCATTTCAAGTTGTC 863
QY 721 gtgagagaaacgggttccgacatgcccgcaacgtggacaggggtctctgcagcttcaag 780
DB 864 GTGAGAGAAACGGCTTCGGACATGCCGCAACGTGGACAGGGTCTCTGACAGCTTCAAG 923
QY 781 atcaatgactcgtgcacactcaatgagaagccctttctgttgaagacacattatttactg 840
DB 924 ATCAATGACTCGGTCCACACTCAATGAGAAAGCCCTTTTCTGTGGAAGATACTTATTACTG 983
QY 841 tctcagcgcttatcttaaaagaagtgtgcatgaaagctgcactccaggtcagcatgaac 900
DB 984 TGTCAGCGCCCTATCTTAAAGAAAGTTGGCATGGAAGCTGCACATCCAGGTGACATGAAC 1043
QY 901 gatggcctctctttatctccagttctgtcatcaccaccacacactgttctgacgt 960
DB 1044 GATGGCTCTCTTTATCTCCAGTTCTGTGTCATCATCACCACCACACACTGTCTGACGGT 1103
QY 961 tccatctggccatcgccctgctgacatctgttctgctgctcctagcctggctctctctg 1020
DB 1104 TCCATCTCTGGCCATCGCCCTGCTGATCTCTGCTCTCTAGCCCTGCTCTCTCTCTG 1163
QY 1021 tggttctggccctctgctgctgactgtgattatcaaggaggtccctccaccctcgcgag 1080
DB 1164 TGGTCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1223
QY 1081 gagagtggagaaaa 1094
DB 1224 GAGAGTGAGGAAGA 1237
RESULT 3
BC012074
LOCUS
DEFINITION Homo sapiens, Similar to tumor endothelial marker 8, clone
MGC:19567 IMAGE:4563020, mRNA, complete cds.
ACCESSION BC012074
VERSION MGC12074.1 GI:15082332
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2112)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DMF
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
```

Series: IRAL Plate: 29 Row: q Column: 3  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 7022737.

FEATURES  
Location/Qualifiers  
1..2112  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:19967 IMAGE:4563020"  
/tissue\_type="Kidney, renal cell adenocarcinoma"  
/clone\_lib="NIH\_MGC\_14"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
113..1114  
/codon\_start=1  
/product="Similar to tumor endothelial marker 8"  
/protein\_id="AAH12074.1"  
/db\_xref="GI:15082333"  
/translation="MATAERRALGIGFQWLSLATLVLCAGGRRDGGPACVGGFD  
LYFLDKSGVLHWNIEIYFVQLAHFISPOLRMSFIVFSTNGITLMKLTEDRQI  
ROGLEQLKVLPGDGYHIEGFERASQIYYENRQGYRTASVIALDGLHEDLFFI  
SERANRDLGALVYGVKDFNQLARADSKHVPVNDGFQALQIIHSLKSC  
SCIETLAEPSTICAGSFQVVRNGFRHARNVRLVCSFKINDSVTLNEKPFVED  
TYLCPDILKEVGMKALQVSMNDGLSFSSSVIITTHCSLHKIASGPTAACME"  
BASE COUNT 573 a 522 c 515 g 502 t  
ORIGIN

Query Match 86.0%; Score 949.4; DB 9; Length 2112;  
Best Local Similarity 99.9%; Pred. No. 2.4e-276;  
Matches 950; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 atggccacggcggagcggagccctcgccatcggtccagttccagtcgtctctttgccaact 60  
DB 113 ATGCCACGCGGAGCGGAGAGCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 172  
QY 61 ctggtgctcatctgcccggcgaaggcagcagcagcagcagcagcagcagcagcagcagcagc 120  
DB 173 CTGGTGCTCATCTGCGCGCGGCAAGGGGACGACGAGGAGTGGGGGTCCAGCCTGCTAC 232  
QY 121 ggcgattgacgtgtactctatttggacaaatccatccggaagtgctgcacacactggaat 180  
DB 233 GCGGATTTGACCTGACTCTTATTTGGCAATCATCAGGAAGTGTGTGCACCACTGGAAT 292  
QY 181 gaaatctatcttggagacagtggtccacaaattcatcagccacagtcagtcagtcagtcag 240  
DB 293 GAAATCTATTACTTTGTGGAACTGTTGGCTCAACAATTATCATCAGCCACACAGTTGAGATG 352  
QY 241 tcctttattgttttccaccgcgaggaacacaccttaataagaaactgcagagaagacagagaa 300  
DB 353 TCCCTTTATTGTTTCTCCACCGGAGGAACACCTTAATGAACACTGACAGAGACAGAGAA 412  
QY 301 caaatcgcgtaaggcctagaaagactccagaagttctgcagagagagacacttacatg 360  
DB 413 CAATTCGCTCAAGGCCATAGAAAGAACTCCAGAAAGTCTGCCAGGAGGACACTTACATG 472  
QY 361 catgaagattgaaaggcgcagtgagcagattattattgaaacacagacagaggtacagg 420  
DB 473 CATGAGGATTTGAAGGGCCAGTGACAGATTTATTATGAAACACAGACAGGATACAGG 532  
QY 421 acagccagcgtcaatctgttgcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 480  
DB 533 ACAGCCAGCGTCATCTGCTTTCACATGATGAGAACTCCCATGAAGATCTCTTTTCTAT 592  
QY 481 tcaagagaggcctcaatagctcgcagatcttggcgaattggtgcaattggttactgtgtgtg 540  
DB 593 TCAGAGAGGGAGGCTAATAGTCTCGAGATCTTGTGCAATTGTTACTGTGTGTGTGTG 652  
QY 541 aaagatttcaatgagacacagctgcccggattgcggacagtaaggatcatgtttcccc 600  
DB 653 AAAGATTTCAATGAGACACAGCTGCGCCCGGATTCGGGACAGTAGAGGATCATGTGTTCCC 712  
QY 601 gtgaatgacggtttccagctctcgaagcgcacatcccaactcaatttgaagaagtcctgc 660

Db 713 GTGAATGACGGCTTTTCAGGCTCTGCAAGGATCATCCACTCAATTTTGAAGAAGTCTCTGC 772  
QY 661 atcgaaattctagcagcgtgaacacatccacatattgtgcagagagtcatttcaagtgtc 720  
Db 773 ATCGAAATCTAGCAGCTGAACCATCCACCATATGTGCAGGAGAGTCATTTCAAGTTGTC 832  
QY 721 gtgagagaaacgcttcgcagatgccccaacgtggacaggggtcctctgcagcgttcaag 780  
Db 833 GTGAGAGGAAACGGCTTCCGACATGCCGCAACGTGGACAGGGTCTCTGCAGCTTCAAG 892  
QY 781 atcaatgactcgtgcacactcaatgagaagccctttctgtgtggaagacacttattactg 840  
Db 893 ATCAATGACTCGGTCACTCAATGAGAAGCCCTTTTCTGTGGAAGATCTTATTACTTG 952  
QY 841 tgcacagcgcctattctaaagaagttggcagaaagcgtgcacccaggtcagcagcatgaac 900  
Db 953 TGTCCAGCGCTATCTTTAAAGAAGTTGGCATGAAGCTGCATCCAGGTCCAGGTGACATGAAC 1012  
QY 901 gatggcctctctttatctccagttctgtcatcatcaccacacacactgt 951  
Db 1013 GATGGCCTCTCTTTATCTCCAGTTCTGTCTCATCATCACCACCACACACTGT 1063  
RESULT 4  
AF378762 5220 bp mRNA linear ROD 07-OCT-2001  
LOCUS Mus musculus tumor endothelial marker 8 precursor (Tem8) mRNA,  
DEFINITION complete cds.  
ACCESSION AF378762  
VERSION AF378762.1 GI:15987504  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 5220)  
AUTHORS Carson-Walter,E.B., Watkins,D.N., Nanda,A., Vogelstein,B.,  
Kinzler,K.W. and St. Croix,B.  
TITLE Cell surface tumor endothelial markers are conserved in mice and  
humans  
JOURNAL Cancer Res. 61 (18), 6649-6655 (2001)  
MEDLINE 21443268  
PUBMED 11595528  
REFERENCE 2 (bases 1 to 5220)  
AUTHORS Carson-Walter,E.B., Vogelstein,B., Kinzler,K.W. and St. Croix,B.  
TITLE Direct Submission  
JOURNAL Submitted (10-MAY-2001) Oncology, Johns Hopkins University, 1650  
Orleans Street, Baltimore, MD 21231, USA  
FEATURES  
Location/Qualifiers  
1..5220  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
1..5220  
/gene="Tem8"  
274..1962  
/gene="Tem8"  
/note="mTEM8"  
/codon\_start=1  
/product="tumor endothelial marker 8 precursor"  
/protein\_id="RAL1199.1"  
/db\_xref="GI:15987505"  
/translation="MDRAGRLGAGLGLCAALVLCAGHGGRRDGGPACVGGFDLY  
FLDKSGVLHWNIEIYFVQLAHFISPOLRMSFIVFSTNGITLMKLTEDRQI  
GLEELQKVLPGDGYHIEGFERASQIYYENRQGYRTASVIALDGLHEDLFFI  
REANRDLGALVYGVKDFNQLARADSKHVPVNDGFQALQIIHSLKSC  
TEFLAEPSTICAGSFQVVRNGFRHARNVRLVCSFKINDSVTLNEKPFVEDTY  
LLCPDILKEVGMKALQVSMNDGLSFSSSVIITTHCSGSIILAILVLLFLLAL  
ALLWFWPLCCVIKEVPPVPEESEDGDLPKKWPVDASVYGGRGVGGIKRM  
EYRWGEKSTEEGAKERAKARVAMPQEEYEPPEPRNLNNMRPSSPRKWSPIK  
KIDALVLLRHGKYDRVSVMRPQDGTGRCTINFRVKNQSPAKYPLNNYHPSSPPAP  
IYTPPPAPPCPPAPAPTPIPSPPTLPPPPQAPPNRRAPPSPRRPPPSV"  
274..354  
/gene="Tem8"  
sig\_peptide

Qy	924	ttctgtcatcatcacaccacacactgttcttgacggtttccactctctgcccactgcccctgct	983
Db	1191	TTCTGTGTCATCATCACACACACTGTTGACAGCGGTCCACTCTCTGGCGATTCGTCTGCT	1250
Qy	984	gatsctgttctcgtctctagccctggctctctctctcgttggttctggccctctgctgcac	1043
Db	1251	GGTCTCTTCTCTGCTGTGGCCCTGGCGCTGCTCTGGTGGTCTGGCCCTCTCTGCTGCAC	1310
Qy	1044	tgtgattacaagagaggtcctccaccctcgcgagagagagtgaggagaaa	1094
Db	1311	AGTGATCATCAAGGAGGTCCCTCCACCCCTGTTGAGGAGAGTGAGGAAGA	1361
RESULT	5		
AK001463			
LOCUS		1436 bp mRNA linear	PRI 22-FEB-2000
DEFINITION		Human sapiens cDNA FLJ10601 fis, clone NT2RP2005000.	
ACCESSION		AK001463	
VERSION		AK001463.1 GI:7022737	
KEYWORDS		oligo capping; fis (full insert sequence).	
SOURCE		Human sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA, clone_lib:NT2RP2 clone:NT2RP2005000.	
ORGANISM		Human sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
AUTHORS		1 (sites)	
		Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K., Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masubo,Y. and Oshima,A.	
TITLE		NEDO human cDNA sequencing project	
JOURNAL		Unpublished (2000)	
REFERENCE		2 (bases 1 to 1436)	
AUTHORS		Isogai,T. and Otsuki,T.	
TITLE		Direct Submission	
JOURNAL		Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)	
COMMENT		NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.	
FEATURES		Location/Qualifiers	
source		1. 1436	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="NT2RP2005000"	
		/cell_line="NT2"	
		/cell_type="teratocarcinoma"	
		/clone_lib="NT2RP2"	
		/note="cloning vector: pME18SFL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction."	
CDS		380..1036	
		/note="unnamed protein product"	
		/codon_start=1	
		/protein_id="BAA91707.1"	
		/db_xref="GI:7022738"	
		/translation="MSFVSTSGTTLMLKLTEDREQIRGLELQVLKLVPCGDTYMHGCFERASEQIYYENKQGYATSVIALTDGELHEDFYSEERANKSRDLGAIYVCVGVDFRQETQARIATADSKDHVPYNDGQFQITLHKSCSTICAGESTICAGESFOVWVRNGFRHARNVDRLVLCSEKINDSVTLRSKLSQSPWVSSTSGFKEGNSHPCLPARPHT"	
BASE COUNT		375 a 349 c 401 g 311 t	
ORIGIN			
Query Match		71.3%	Score 787; DB 9; Length 1436.

Best Local Similarity 99.3%; Pred. No. 4e-227; Matches 801; Conservative 0; Mismatches 5; Indels 1; Gaps 1;									
QY	1	atggccacaggcggagcggagagccctcgccatcggttccagtggtctctttggccact	60						
Db	144	ATGGCCACGGCGGAGCGGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTC-ACGGCCACT	202						
QY	61	ctggtgctcatctgcgcggcgcaagggggacgcaggagggatgggggtccacgtgctac	120						
Db	203	CTGGTGTCTCATCTGCGCGGCAAGGGGACGCAGGAGGATGGGGTCCAGCTGCTAC	262						
QY	121	ggcggatttgacctgtactctattttggacaatacagggaagtgtcgtgcaccactggaat	180						
Db	263	GGCGGATTTGACCTGTACTTCAATTTTGGCAATCAGGAAGTGTGTGCACCACTGGAT	322						
QY	181	gaattctattcttggaaactgtggtccaaaattcatcagccacagttgagaatg	240						
Db	323	GAATCTATTACTTTGTGGACAGCTTGGCTCACAAATTCATCAGCCACAGTTGAGAATG	382						
QY	241	tcctttattgtttctccaccgccaggaaacaccttaataaactgcacagaacagagaa	300						
Db	383	TCCTTTATTGTTTCTCACCCGAGGAACAACCTTAATGAAACTGACAGAAGACAGAA	442						
QY	301	caaatcgtcaagccttagaagaactccagaagtcttcaggagagacacttacatg	360						
Db	443	CAAAATCGCTCAAGGCCCTAGAAAGAACTCCAGAAAGTCTGCCAGGAGGACACTTACATG	502						
QY	361	catgaagatttgaaggccagtcagagcagatttattatgaaacagacagaaggtacagg	420						
Db	503	CATGAAGGATTTGAAAGGCCAGTGAGCAGATTATTATGAAACACACAAGGGTACAGG	562						
QY	421	acagccagcgtcatcattgttctaactgatggagaactccatgaagatctcttttctat	480						
Db	563	ACAGCTAGCGTCATCTGCTTTGACTGTATGGAACTCCATGAAGATCTCTTTTCTAT	622						
QY	481	tcagagaggaggtcaatagctctcgagatcttggtgcaattgtttactgtgtggtg	540						
Db	623	TCAGAGAGGAGGCTAATAGCTCTCGAGATCTTGTGCAATTTGTTACTGTGTGGTGTG	682						
QY	541	aaagatttcaatgagacacagctggccgagttcgcagactgaagcatcgtgtttccc	600						
Db	683	AAAGATTTCAATGAGACACAGCTGCGCCGGATTGCGGACAGTAAGGATCATGTGTTCCC	742						
QY	601	gtgaatgacggttttcaggctctcgaaggcatcatccactcaattttgaagaagctctgc	660						
Db	743	GTGAATGACGGCTTTTCAGGCTCTGCAAGGCATCATCCACTCAAATTTTGAAGAAGTCTCGC	802						
QY	661	atcgaattcttagcagctgaaccatccaccatatgtgcaggagagtcatttcaagtgtc	720						
Db	803	ATCGAAATTTTAGCAGCTGAACCATCCACCATATGTGCAGGAGAGTCATTTCAAAGTTGTC	862						
QY	721	gtgagaggaaacggtctccgacatgcccgcaactggcagaggtcctctgcagettcaag	780						
Db	863	GTGAGAGGAAACGGCTTCGACATGCCCGCAACGTGGACAGGGTCCCTGCGAGTTCAAG	922						
QY	781	atcaatgactcgggtcacactcaatgag	807						
Db	923	ATCAATGACTCGGTCACTCAGTAAG	949						
RESULT	6								
LOCUS	AK025429								
DEFINITION	Homo sapiens cDNA: FLJ21776 fis, clone HEP00171.								
ACCESSION	AK025429								
VERSION	AK025429.1								
KEYWORDS	oligo capping; fis (full insert sequence);								
SOURCE	Homo sapiens hepatoma cell_line:HepG2 cDNA to mRNA, clone_lib:HEP								
ORGANISM	Homo sapiens								
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
REFERENCE	1 (sites)								
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okikani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S. NEDO human cDNA sequencing project									
TITLE Unpublished (2000)									
JOURNAL 2 (bases 1 to 4007)									
REFERENCE Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y. Direct Submission									
AUTHORS Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdnaife@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)									
COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).									
FEATURES									
source									
1..4007									
/organism="Homo sapiens"									
/db_xref="taxon:9606"									
/clone="HEP00171"									
/cell_line="HepG2"									
/cell_type="hepatoma"									
/clone_lib="HEP"									
/note="cloning vector pME18SFL3"									
323..1147									
/note="unnamed protein product"									
/codon_start=1									
/protein_id="BAB15128.1"									
/db_xref="GI:10437939"									
/translation="MKAALQVSMNDGLSFSSVITTHCSGDSILAIALLIFLLL									
ALALLWFWPLCCVTWIKVPPPPPAEAESEEDDGLPKKWPVTVDASVYGGVGGIK									
RNEVKGKSGTEEGAKLEKAKARVKMPEQVEFEPENLNNMRRSSPRKWTSP									
KQKLDALWLLRKGYDRVSMRPQDGTGRCINFRVKNQPKAPYPLANAVHTSSPPP									
APIYTPPPAPHCPPPPSPAPTPPISPPSTLPPPPAPPNRPSPRPPSPV"									
BASE COUNT	1144	a	1003	c	832	g	1028	t	
ORIGIN									
Query Match 49.2%; Score 542.8; DB 9; Length 4007;									
Best Local Similarity 99.6%; Pred. No. 6.2e-153;									
Matches 544; Conservative 0; Mismatches 2; Indels 0; Gaps 0;									
QY	549	caatgagacacagctggccggattgcggacagtgaaggatcatgtgtttcccgatgaatga	608						
Db	1	CAATGAGACACAGCTGGCCGGATTGCGGACAGTAAGGATCATGTGTTCCCGTGAATGA	60						
QY	609	cggctttcaggctctgcaaggcatctccactcaattttgaagaagctcctgcacgaat	668						
Db	61	CGGCTTTACGGCTCTGCAAGGCATCATCCACTCAATTTTGAAGAAGTCTCTGCATCGAAAT	120						
QY	669	tctagcagctgaaccatccaccatatgtcagagagatcatttcaagttcgtgagagg	728						
Db	121	TCTAGCAGCTGAACCATCCACCATATGTGCAGGAGAGTCATTTCAGTTCCTGTGAGAG	180						
QY	729	aaacggcttccgacatgcccgcaacgtggcagggttccctctgcagcttcaagatcaatga	788						
Db	181	AAACGGCTTCGACATGCCCGCAACGTGGACAGGGTCTCTTCGAGCTTCAAGATCAATGA	240						
QY	789	ctcgggtcacactcaatgagaagccctttctgtggaagacacttatttactgtgtccagc	848						
Db	241	CTCGGTCACTCAATGAGAAGCCCTTTCTGTGGAAGATACATTATTACTGTGTCCAGC	300						
QY	849	gcctatcttaaaagaagtgtgcatgaaagctgcactccaggtcagcatgaacgatggcct	908						
Db	301	GCCATCTTAAAAAGAGTTGGCATGAAAGCTGCACCTCCAGGTCAAGTCAAGTGCCT	360						
QY	909	ctcttttatctccagttctgtcatcatcaccacacactgttctgacggttccatcct	968						



REMARK COMMENT  
NIH-MGC Project URL: http://mgc.nci.nih.gov  
Contact: MGC help desk  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: DCTD/DTP  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: http://www.hgsc.bcm.tmc.edu/cdna/  
Contact: villalob@bcm.tmc.edu.  
Villalón, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Series: IRAK Plate: 28 Row: k Column: 23  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 10437939.

FEATURES  
source  
1..1297  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4705862"  
/tissue\_type="Breast, mammary adenocarcinoma."  
/clone\_lib="NIH\_MGC\_87"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
<1..295  
/codon\_start=2  
/product="Unknown (protein for IMAGE:4705862)"  
/protein\_id="AAH12475.1"  
/db\_xref="GI:15214684"  
/translation="FOVVVRNGFRHARNVDSFLGSKINDSVTLNEKPFVVEDTYLL  
CPAPILKEVGKMAALQVSMNDGLSFISSVIITTTQGSILHKIASGPTTAACME"

BASE COUNT 365 a 333 c 276 g 323 t  
ORIGIN  
Query Match 21.4%; Score 236; DB 9; Length 1297;  
Best Local Similarity 98.0%; Pred. No. 5.5e-60;  
Matches 239; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 708 attcaagttgtgtgagagaaacggctccgacatgcccgcaacgtgacaggtcct 767  
|||||  
Db 1 ATTTCAAGTTGTCGTGAGAGAAACGGCTTCCGACATGCCCCGCAACGTGGACAGCTTCCT 60  
Qy 768 ctgcagcttcaagatcaatgactcgtgcacactcaatgagaagccctttctctggaaga 827  
|||||  
Db 61 CTGCAGCTTCAAGATCAATGACTCGGTGCACATCAATGAGAAGCCCTTTCTGTGAAGA 120  
Qy 828 cacttattactgtgtcagcgccctatcttaaaagaagttggcaatgaactgaactcca 887  
|||||  
Db 121 TACTTATTTACTGTGTCCAGCGCCTATCTTAAAAAGAGTTGGCATGAAGCTGCACATCCA 180  
Qy 888 ggtcagatgaacgatggcctcttttattctccagttctgtcatcatcaccacacaca 947  
|||||  
Db 181 GGTCAAGATGAACGATGCGCTCTCTTTATCTCCAGTTCGTGCATCATCACCACACACA 240  
Qy 948 ctgt 951  
||  
Db 241 AGGT 244

RESULT 9  
AK055636 2026 bp mRNA linear PRI 31-OCT-2001  
LOCUS  
DEFINITION Homo sapiens cDNA FLJ31074 fis, clone HSYRA2001476.  
ACCESSION AK055636

VERSION  
KEYWORDS  
SOURCE

AK055636.1 GI:16550414  
oligo capping; fis (full insert sequence).  
Homo sapiens synovioocytes from rheumatoid arthritis (HS-RA) cDNA  
to mRNA, clone\_lib:HSYRA2 clone:HSYRA2001476.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS  
1 (sites)  
Ninomiya, K., Wagatsuma, M., Kanda, K., Kondo, H., Yokoi, T.,  
Kodaira, H., Furuya, T., Takahashi, M., Kikkawa, E., Omura, Y., Abe, K.,  
Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M.,  
Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S.,  
Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T.,  
Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M.,  
Kikuchi, H., Murakawa, K., Kanohori, K., Takahashi-Fujii, A.,  
Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,  
Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.  
NEDO human cDNA sequencing project

TITLE

Unpublished

JOURNAL

2 (bases 1 to 2026)

REFERENCE

Isogai, T., Otsuki, T. and Sugiyama, T.

AUTHORS

Direct Submission

TITLE

Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,

JOURNAL

Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

COMMENT

(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3951, Fax: 81-438-52-3952)  
NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: RAB and  
HRI.

FEATURES

Location/Qualifiers

source

1..2026  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HSYRA2001476"  
/cell\_type="synovioocytes from rheumatoid arthritis  
(HS-RA)"  
/clone\_lib="HSYRA2"  
/note="cloning vector: pME18SFL3-primary culture,  
synovioocytes from rheumatoid arthritis"  
46..783  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB70976.1"  
/db\_xref="GI:16550415"  
/translation="MRLSFIVFSSQATIIPLTCDRGKISKGLDLKRVSPVGETYIH  
EGLKLANEQIKAGGLKTSIIIALTDGKLDGLVPSYAEKAKISRLGASVYGVFL  
DPEQAQLRIADSKEQVFPVKGFQALKIINSILAOISCTEILELPSSVCYGEFVI  
VLSGRGFMGSRNGSVLCTVNETYTSVKRPVSVQLNSMLCPAPILINKAGEWGLTVT  
QAGVKWHDLTHTCFLGSGSDPPTSAS"

CDS

BASE COUNT 583 a 401 c 502 g 540 t  
ORIGIN  
Query Match 21.0%; Score 232.2; DB 9; Length 2026;  
Best Local Similarity 59.9%; Pred. No. 8.7e-59;  
Matches 409; Conservative 0; Mismatches 268; Indels 6; Gaps 1;  
Qy 187 tattactttgtggagacagttggctcacaattcatcagccacagttgagaatgctctt 246  
|||||  
Db 1 TATAATTTGCTACAGCAACTTGGGAGAGATTTGTGAGCCCCGAAATGAGATTATCTTTC 60  
Qy 247 attgtttttctccaccggaggaacaaccttaataatgaactgacagagagagacaaatc 306  
|||||  
Db 61 ATTGTGTTTTCTTCTCAAGCAACTATTATTTTGGCATTAACTGGAGAGAGCAAAATC 120  
Qy 307 cgtcaaggccttagaagaactcccaagagttctgcagagagagacactcacatgcatgaa 366  
|||||  
Db 121 AGTAAAGGCTTGAGGAGTTTAAACCGTGTAGTCCAGGAGAGACATATATCCATGAA 180







```

SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   Eukaryota: Chordata; Craniata: vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
TITLE       1 (bases 1 to 81017)
JOURNAL     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
REFERENCE   Homo sapiens chromosome 4, clone RP11-85D17
AUTHORS     2 (bases 1 to 81017)
            Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bedalov,F.,
            Boguski,M.S., Bouckhaghter,B., Brown,A., Burkett,G.,
            Campione,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collamore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehotzky,J.,
            Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneses,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
            Direct Submission
TITLE       Submitted (03-MAR-2000) Whitehead Institute/MIT Center for Genome
JOURNAL     Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIGR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L7756
            Center clone name: 85_D_17
            -----
            * NOTE: This record contains 94 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *
            * 1 709: contig of 709 bp in length
            * 710 809: gap of 100 bp
            * 810 1567: contig of 758 bp in length
            * 1568 1667: gap of 100 bp
            * 1668 2433: contig of 766 bp in length
            * 2434 2533: gap of 100 bp
            * 2534 3310: contig of 777 bp in length
            * 3311 3410: gap of 100 bp
            * 3411 4176: contig of 766 bp in length
            * 4177 4276: gap of 100 bp
            * 4277 5036: contig of 760 bp in length
            * 5037 5136: gap of 100 bp
            * 5137 5899: contig of 763 bp in length
            * 5900 5999: gap of 100 bp
            * 6000 6764: contig of 765 bp in length
            * 6765 6864: gap of 100 bp
            * 6865 7612: contig of 748 bp in length
            *
            * 7613 7712: gap of 100 bp
            * 7713 8467: contig of 755 bp in length
            * 8468 8567: gap of 100 bp
            * 8568 9320: contig of 753 bp in length
            * 9321 9420: gap of 100 bp
            * 9421 10187: contig of 767 bp in length
            * 10188 10287: gap of 100 bp
            * 10288 11042: contig of 755 bp in length
            * 11043 11142: gap of 100 bp
            * 11143 11910: contig of 788 bp in length
            * 11911 12010: gap of 100 bp
            * 12011 12757: contig of 747 bp in length
            * 12758 12857: gap of 100 bp
            * 12859 13619: contig of 762 bp in length
            * 13620 13719: gap of 100 bp
            * 13720 14493: contig of 774 bp in length
            * 14494 14593: gap of 100 bp
            * 14594 15348: contig of 755 bp in length
            * 15349 15448: gap of 100 bp
            * 15449 16206: contig of 758 bp in length
            * 16207 16306: gap of 100 bp
            * 16307 17054: contig of 748 bp in length
            * 17055 17154: gap of 100 bp
            * 17155 17900: contig of 746 bp in length
            * 17901 18000: gap of 100 bp
            * 18001 18767: contig of 767 bp in length
            * 18768 18867: gap of 100 bp
            * 18868 19642: contig of 775 bp in length
            * 19643 19742: gap of 100 bp
            * 19743 20493: contig of 751 bp in length
            * 20494 20593: gap of 100 bp
            * 20594 21337: contig of 744 bp in length
            * 21338 21437: gap of 100 bp
            * 21438 22202: contig of 765 bp in length
            * 22203 22302: gap of 100 bp
            * 22303 23072: contig of 770 bp in length
            * 23073 23172: gap of 100 bp
            * 23173 23941: contig of 769 bp in length
            * 23942 24041: gap of 100 bp
            * 24042 24799: contig of 758 bp in length
            * 24800 24899: gap of 100 bp
            * 24900 25667: contig of 768 bp in length
            * 25668 25767: gap of 100 bp
            * 25768 26515: contig of 748 bp in length
            * 26516 26615: gap of 100 bp
            * 26616 27375: contig of 760 bp in length
            * 27376 27475: gap of 100 bp
            * 27476 28238: contig of 763 bp in length
            * 28239 28338: gap of 100 bp
            * 28339 29104: contig of 766 bp in length
            * 29105 29204: gap of 100 bp
            * 29205 29969: contig of 765 bp in length
            * 29970 30069: gap of 100 bp
            * 30070 30837: contig of 768 bp in length
            * 30838 30937: gap of 100 bp
            * 30938 31710: contig of 773 bp in length
            * 31711 31810: gap of 100 bp
            * 31811 32576: contig of 766 bp in length
            * 32577 32676: gap of 100 bp
            * 32677 33461: contig of 785 bp in length
            * 33462 33561: gap of 100 bp
            * 33562 34337: contig of 776 bp in length
            * 34338 34437: gap of 100 bp
            * 34438 35216: contig of 779 bp in length
            * 35217 35316: gap of 100 bp
            * 35317 36082: contig of 766 bp in length
            * 36083 36182: gap of 100 bp
            * 36183 36951: contig of 769 bp in length
            * 36952 37051: gap of 100 bp
            * 37052 37815: contig of 764 bp in length
            * 37816 37915: gap of 100 bp
            * 37916 38676: contig of 761 bp in length
            * 38677 38776: gap of 100 bp
            *

```

```

* 38777 39563: contig of 787 bp in length
* 39564 39663: gap of 100 bp
* 39664 40445: contig of 782 bp in length
* 40446 40545: gap of 100 bp
* 40546 41316: contig of 771 bp in length
* 41317 41416: gap of 100 bp
* 41417 42181: contig of 765 bp in length
* 42182 42281: gap of 100 bp
* 42282 43030: contig of 749 bp in length
* 43031 43130: gap of 100 bp
* 43131 43913: contig of 783 bp in length
* 43914 44013: gap of 100 bp
* 44014 44770: contig of 757 bp in length
* 44771 44870: gap of 100 bp
* 44871 45631: contig of 761 bp in length
* 45632 45731: gap of 100 bp
* 45732 46487: contig of 756 bp in length
* 46488 46587: gap of 100 bp
* 46588 47340: contig of 753 bp in length
* 47341 47440: gap of 100 bp
* 47441 48171: contig of 731 bp in length
* 48172 48271: gap of 100 bp
* 48272 49028: contig of 757 bp in length
* 49029 49128: gap of 100 bp
* 49129 49883: contig of 755 bp in length
* 49884 49983: gap of 100 bp
* 49984 50748: contig of 765 bp in length
* 50749 50848: gap of 100 bp
* 50849 51619: contig of 771 bp in length
* 51620 51719: gap of 100 bp
* 51720 52473: contig of 754 bp in length
* 52474 52573: gap of 100 bp
* 52574 53327: contig of 754 bp in length
* 53328 53427: gap of 100 bp
* 53428 54198: contig of 771 bp in length
* 54199 54298: gap of 100 bp
* 54299 55054: contig of 756 bp in length
* 55055 55154: gap of 100 bp
* 55155 55921: contig of 767 bp in length
* 55922 56021: gap of 100 bp
* 56022 56781: contig of 760 bp in length
* 56782 56881: gap of 100 bp
* 56882 57638: contig of 757 bp in length
* 57639 57738: gap of 100 bp
* 57739 58492: contig of 754 bp in length
* 58493 58592: gap of 100 bp
* 58593 59364: contig of 772 bp in length
* 59365 59464: gap of 100 bp

Query Match 6.5%; Score 72.2; DB 2; Length 81017;
Best Local Similarity 96.1%; Pred. No. 7.6e-10;
Matches 74; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 803 atgagaagccctttctgtggaagacacattattactgtgtccagcgctattcttaaaag 862
|||||
Db 74775 ATGAGAAGCCCTTTCTGTGGAAGATACATTATTACTGTGTCCAGCGCCTTAATAAG 74834
|||||

Qy 863 aagttggcatgaaagct 879
|||||
Db 74835 AAGTTGCATGTAAGTT 74851
|||||

RESULT 14
HSM800800
LOCUS
DEFINITION Homo sapiens mRNA; cDNA DKF2p586F1324 (from clone DKF2p586F1324).
ACCESSION AL110155
VERSION AL110155.1 GI:5817057
KEYWORDS human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1877 bp mRNA linear PRI 18-FEB-2000
HSM800800
DEFINITION Homo sapiens mRNA; cDNA DKF2p586F1324 (from clone DKF2p586F1324).
ACCESSION AL110155
VERSION AL110155.1 GI:5817057
KEYWORDS human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```

1 (bases 1 to 1877)
Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and
Wiemann,S.
Direct Submission
Submitted (15-AUG-1999) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKF2p586F1324) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.

FEATURES
Location/Qualifiers
1..1877
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKF2p586F1324"
/cloned_lib="586 (synonym: hutel). Vector pSport1; host
DH10B; sites NotI + SalI/MLui"
/dev_stage="adult"
/tissue_type="uterus"
polyA_signal 1840..1845
polyA_site 1858
BASE COUNT 525 a 421 c 495 g 436 t
ORIGIN

Query Match 4.3%; Score 48; DB 9; Length 1877;
Best Local Similarity 79.2%; Pred. No. 0.0064;
Matches 57; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1026 ctggccctctgtgcactgtgtattatcaaggagggtccctccaccctgcagagagag 1085
|||||
Db 1513 CTGTCCTCTGTTTCTTTTCAGATTATCAAGAGGTCCTCCACCCCTCCCGAGAGAG 1572
|||||

Qy 1086 tgaggaataataa 1097
|||||
Db 1573 TGAGGTAAGTGA 1584
|||||

RESULT 15
AC025010/c
LOCUS
DEFINITION Homo sapiens chromosome 4 clone RP11-85D17 map 4, linear HTG 13-JUL-2000
AC025010
VERSION AC025010.1 GI:7145054
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 81017)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 4, clone RP11-85D17
Unpublished
2 (bases 1 to 81017)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgaiter,B., Brown,A., Burkett,G.,
Campopiano,A., Castelle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McKernan,K., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
```

Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

## Direct Submission

Submitted (03-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L7756

Center clone name: 85\_D\_17

-----

\* NOTE: This record contains 94 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 709: contig of 709 bp in length

\* 710 809: gap of 100 bp

\* 810 1567: contig of 758 bp in length

\* 1568 1667: gap of 100 bp

\* 1668 2433: contig of 766 bp in length

\* 2434 2533: gap of 100 bp

\* 2534 3310: contig of 777 bp in length

\* 3311 3410: gap of 100 bp

\* 3411 4176: contig of 766 bp in length

\* 4177 4276: gap of 100 bp

\* 4277 5036: contig of 760 bp in length

\* 5037 5136: gap of 100 bp

\* 5137 5899: contig of 763 bp in length

\* 5900 5999: gap of 100 bp

\* 6000 6764: contig of 765 bp in length

\* 6765 6864: gap of 100 bp

\* 6865 7612: contig of 748 bp in length

\* 7613 7712: gap of 100 bp

\* 7713 8467: contig of 755 bp in length

\* 8468 8567: gap of 100 bp

\* 8568 9320: contig of 753 bp in length

\* 9321 9420: gap of 100 bp

\* 9421 10187: contig of 767 bp in length

\* 10188 10287: gap of 100 bp

\* 10288 11042: contig of 755 bp in length

\* 11043 11142: gap of 100 bp

\* 11143 11910: contig of 768 bp in length

\* 11911 12010: gap of 100 bp

\* 12011 12757: contig of 747 bp in length

\* 12758 12857: gap of 100 bp

\* 12858 13619: contig of 762 bp in length

\* 13620 13719: gap of 100 bp

\* 13720 14493: contig of 774 bp in length

\* 14494 14593: gap of 100 bp

\* 14594 15348: contig of 755 bp in length

\* 15349 15448: gap of 100 bp

\* 15449 16206: contig of 758 bp in length

\* 16207 16306: gap of 100 bp

\* 16307 17054: contig of 748 bp in length

\* 17055 17154: gap of 100 bp

\* 17155 17900: contig of 746 bp in length

\* 17901 18000: gap of 100 bp

\* 18001 18767: contig of 767 bp in length

\* 18768 18867: gap of 100 bp

\* 18868 19642: contig of 775 bp in length

\* 19643 19742: gap of 100 bp

\* 19743 20493: contig of 751 bp in length

\* 20494 20593: gap of 100 bp

\* 20594 21333: contig of 744 bp in length

\* 21338 21437: gap of 100 bp

\* 21438 22202: contig of 765 bp in length

\* 22203 22302: gap of 100 bp

\* 22303 23072: contig of 770 bp in length

\* 23073 23172: gap of 100 bp

\* 23173 23941: contig of 769 bp in length

\* 23942 24041: gap of 100 bp

\* 24042 24799: contig of 758 bp in length

\* 24800 24899: gap of 100 bp

\* 24900 25667: contig of 768 bp in length

\* 25668 25767: gap of 100 bp

\* 25768 26515: contig of 748 bp in length

\* 26516 26615: gap of 100 bp

\* 26616 27375: contig of 760 bp in length

\* 27376 27475: gap of 100 bp

\* 27476 28238: contig of 763 bp in length

\* 28239 28338: gap of 100 bp

\* 28339 29104: contig of 766 bp in length

\* 29105 29204: gap of 100 bp

\* 29205 29969: contig of 765 bp in length

\* 29970 30069: gap of 100 bp

\* 30070 30837: contig of 768 bp in length

\* 30838 30937: gap of 100 bp

\* 30938 31710: contig of 773 bp in length

\* 31711 31810: gap of 100 bp

\* 31811 32576: contig of 766 bp in length

\* 32577 32676: gap of 100 bp

\* 32677 33461: contig of 785 bp in length

\* 33462 33561: gap of 100 bp

\* 33562 34337: contig of 776 bp in length

\* 34338 34437: gap of 100 bp

\* 34438 35216: contig of 779 bp in length

\* 35217 35316: gap of 100 bp

\* 35317 36082: contig of 766 bp in length

\* 36083 36182: gap of 100 bp

\* 36183 36951: contig of 769 bp in length

\* 36952 37051: gap of 100 bp

\* 37052 37815: contig of 764 bp in length

\* 37816 37915: gap of 100 bp

\* 37916 38676: contig of 761 bp in length

\* 38677 38776: gap of 100 bp

\* 38777 39563: contig of 787 bp in length

\* 39564 39663: gap of 100 bp

\* 39664 40445: contig of 782 bp in length

\* 40446 40545: gap of 100 bp

\* 40546 41316: contig of 771 bp in length

\* 41317 41416: gap of 100 bp

\* 41417 42181: contig of 765 bp in length

\* 42182 42281: gap of 100 bp

\* 42282 43030: contig of 749 bp in length

\* 43031 43130: gap of 100 bp

\* 43131 43913: contig of 783 bp in length

\* 43914 44013: gap of 100 bp

\* 44014 44770: contig of 757 bp in length

\* 44771 44870: gap of 100 bp

\* 44871 45631: contig of 761 bp in length

\* 45632 45731: gap of 100 bp

\* 45732 46487: contig of 756 bp in length

\* 46488 46587: gap of 100 bp

\* 46588 47340: contig of 753 bp in length

\* 47341 47440: gap of 100 bp

\* 47441 48171: contig of 731 bp in length

\* 48172 48271: gap of 100 bp







Claim 11; Page 505-506; 562pp; English.

PS AAD05300-AAD05379 represent cDNAs corresponding to 28 human secreted  
 XX protein genes, and AAE01436-AAE01513 represent the proteins they encode.  
 CC The genes and their secreted human secreted protein fragments or variants.  
 CC AAE01514-AAE01544 represent human secreted protein fragments or variants.  
 CC The genes and their secreted proteins are useful for preventing,  
 CC treating or ameliorating medical conditions, e.g., by protein or gene  
 CC therapy. Pathological conditions can be diagnosed by determining the  
 CC amount of the new protein in a sample or by determining the presence of  
 CC mutations in the new genes. Specific uses are described for each of the  
 CC 28 genes, based on the tissues in which they are most highly expressed,  
 CC and include developing products for the diagnosis and treatment of  
 CC proliferative disorders, cancer, tumours, foetal and developmental  
 CC abnormalities, haematopoietic disorders, diseases of the immune system,  
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
 CC allergies, neurological disorders (e.g., Alzheimer's disease,  
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,  
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
 CC disorders, and infectious. The proteins can also be used to aid wound  
 CC healing and epithelial cell proliferation, to prevent skin aging due to  
 CC sunburn, to maintain organs before transplantation, for supporting cell  
 CC culture of primary tissues, to regenerate tissues, to identify their  
 CC cognate ligands or binding partners, and in chemotaxis, and can be used  
 CC as a food additive or preservative to modify storage properties.  
 CC Antibodies specific for a protein of the invention can be used in  
 CC alleviating symptoms associated with the disorders mentioned above, and  
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
 CC immunosorbent assay (ELISA). The present sequence represents a human  
 CC secreted protein of the invention.  
 XX  
 SQ Sequence 403 AA;

Query Match 97.7%; Score 1870; DB 22; Length 403;  
 Best Local Similarity 99.2%; Pred. No. 4.6e-191;  
 Matches 360; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 MATAERRALGIGFOWLSLTLVLICAGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60  
 Db 1 mataerralgigfowlsatlvlilicaggrrdggpacyggfdlyfldksgsvlhhwn 60  
 QY 61 EIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMLKLTEDREQIROGLEELQKVLPGGDTYM 120  
 Db 61 eiylfveqlahkfispolrmsfivfstrgttlmlkltedreqirgleelqkvlpggdtym 120  
 QY 121 HEGFERASEQIYYENROGYRTASVITIALTDGELHEDLFFYSERANRSDIGAIVYCVGV 180  
 Db 121 hegferaseqiyyenrgyrtasvialtdgelhedlffysereanrsrdigaivycvgv 180  
 QY 181 KDFNETOLARIADSKDHVPVNDGQFALOGTIHSILKSCIEILAAEPSTICAGESFQWV 240  
 Db 181 kdfnetolariadskdhvfvndgqfalqgltihsilksclieilaepsticagesfqv 240  
 QY 241 VRGNFGRHARNVDRVLCSEFKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300  
 Db 241 vrgnfgfharndrvlcsfkindsvtlnekpfsvedtyllcpapilkevgmkaalqvsmn 300  
 QY 301 DGLSFISSVIITTHCSGSLIAUILLPLLLALALLWWFPLCCTVIKEVPPPPAE 360  
 Db 301 dglsfissviiitthcsdgsilailllplllalallwwfplcctviikevppppae 360  
 QY 361 ESE 363  
 Db 361 ese 363  
 RESULT 3  
 AAB01422  
 ID AAB01422 standard; Protein; 333 AA.  
 XX  
 AC AAB01422;

XX 20-OCT-2000 (first entry)  
 DT Human TANGO 197.  
 XX  
 DE TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;  
 KW graft versus-host diseases; rheumatoid arthritis; psoriasis;  
 KW inflammatory bowel disease; septic shock; ulcerative colitis;  
 KW Crohn's disease; chronic myelogenous leukemia; cancer; liver  
 KW disease; Hodgkin's disease; osteoarthritis; Lyme's disease;  
 KW cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;  
 KW systemic lupus erythematosus; transgenic animal; diagnosis;  
 KW prognosis; prophylactic; therapeutic; human.  
 XX Homo sapiens.  
 OS  
 XX WO2000039284-A1.  
 PN 06-JUL-2000.  
 XX  
 PD 23-DEC-1999; 99WO-US31025.  
 PF  
 XX 30-DEC-1998; 98US-0223546.  
 PR  
 XX (MILL-) MILLENNIUM PHARM INC.  
 PA Holtzman DA;  
 PI  
 XX WPI; 2000-465743/40.  
 DR N-PSDB; AAA47455.  
 XX  
 XX Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,  
 PT 224 and 239 polypeptides useful for the treatment of asthma, rheumatoid  
 PT arthritis, psoriasis and autoimmune diseases  
 PT  
 XX Claim 8; Fig 4; 209pp; English.  
 PS  
 XX Nucleic acids encoding TANGO polypeptides are useful as modulating  
 CC agents for regulating cellular processes like asthma, graft  
 CC versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory  
 CC bowel disease, septic shock, ulcerative colitis, Crohn's disease,  
 CC chronic myelogenous leukemia, cancer, liver disease, Hodgkin's  
 CC disease, osteoarthritis, Lyme's disease, cachexia and autoimmune  
 CC diseases e.g. myasthenia gravis, autoimmune diabetes and systemic  
 CC lupus erythematosus. The nucleic acids are also useful for producing  
 CC transgenic animals and the TANGO polypeptides themselves. Partial  
 CC TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in  
 CC forensic biology, for diagnostic assays, prognostic assays,  
 CC pharmacogenomics and for monitoring clinical trials. TANGO  
 CC polypeptides are suitable for both prophylactic and therapeutic  
 CC methods for treating a subject at risk of a disorder or having a  
 CC disorder associated with aberrant TANGO expression. A wide range  
 CC of cellular disorders can be treated.  
 XX  
 SQ Sequence 333 AA;  
 Query Match 86.2%; Score 1649; DB 21; Length 333;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-167;  
 Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MATAERRALGIGFOWLSLTLVLICAGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60  
 Db 1 mataerralgigfowlsatlvlilicaggrrdggpacyggfdlyfldksgsvlhhwn 60  
 QY 61 EIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMLKLTEDREQIROGLEELQKVLPGGDTYM 120  
 Db 61 eiylfveqlahkfispolrmsfivfstrgttlmlkltedreqirgleelqkvlpggdtym 120  
 QY 121 HEGFERASEQIYYENROGYRTASVITIALTDGELHEDLFFYSERANRSDIGAIVYCVGV 180  
 Db 121 hegferaseqiyyenrgyrtasvialtdgelhedlffysereanrsrdigaivycvgv 180

Fri Aug 9 10:56:58 2002

us-09-970-076-2.rag

QY 181 KDFNETQLARIADSKDHVPVNDGFOALOGIIHSILKSKSCIELAAEPSTICAGESFQVW 240  
|||||  
Db 181 kdfnetqlariadskdhvpvndgfgalqgiihsilkskscielaaepsticagesfqv 240  
  
QY 241 VRNGGFRHARNVDRVLCSEKINDSVTLNKPFSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
|||||  
Db 241 vrgngfrharnvdrvlcsfkindsvtlnkpfsvedtyllcpapilkevgmkaalqvsmn 300  
  
QY 301 DGLSFSSSSVITTHCS 318  
|||||  
Db 301 dglssfvssvittthcs 318  
  
RESULT 4  
AAM38976  
ID AAM38976 standard; Protein; 297 AA.  
XX  
AC AAM38976;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 2121.  
XX  
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
OS Homo sapiens.  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US34263.  
XX  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX  
WPI; 2001-442253/47.  
DR N-PSDB; AAI58132.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
PS Example 4; SEQ ID NO 2121; 10078pp; English.  
XX  
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic.  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 297 AA;  
  
Query Match 72.7%; Score 1392; DB 22; Length 297;  
Best Local Similarity 99.3%; pred. No. 4e-140;  
Matches 267; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MATAERRALGIGFQWLSLATVILICAGGRRDEGGPACYGFDLYFLDKSGSVLHHWN 60  
|||||  
Db 1 mataerralgigfqlslatvlicagggrrredgppacygfdlyfldksqsvlhhwn 60  
  
QY 61 ELYYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDGYM 120  
|||||  
Db 61 elyyfveqlahkfispqlrmsfivstrgttlmkltedreqirggleelqkvlpggdtym 120  
  
QY 121 HEGFERASEQIYYENRQGYRTASVITIALTDGELHEDLFFYSERANRSRDIGALVYCVGV 180  
|||||  
Db 121 hegferaseqiyyenrqqyrtasvialtdgelhedlffysereanrsrdigalvycgv 180  
  
QY 181 KDFNETQLARIADSKDHVPVNDGFOALOGIIHSILKSKSCIELAAEPSTICAGESFQVW 240  
|||||  
Db 181 kdfnetqlariadskdhvpvndgfgalqgiihsilkskscielaaepsticagesfqv 240  
  
QY 241 VRNGGFRHARNVDRVLCSEKINDSVTLNE 269  
|||||  
Db 241 vrgngfrharnvdrvlcsfkindsvtlsk 269  
  
RESULT 5  
AAB92985  
ID AAB92985 standard; Protein; 218 AA.  
XX  
AC AAB92985;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:11706.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
PS Claim 8; SEQ ID 11706; 2537pp + CD ROM; English.  
XX



CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX Sequence 218 AA;

Query Match 50.5%; Score 966; DB 22; Length 218;  
 Best Local Similarity 98.9%; Pred. No. 9.5e-95;  
 Matches 186; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 80 MSFVSTRGTTMLKLTEDRQIRQGLELQKVLPGDVTYMHGFERASEQIYYENRQGY 139  
 Db 1 msfivstrgttlmkltdreqirggleelqkvlpvgdtymhgferaseqiyenrqqy 60

Qy 140 RTASVITALTGDELHEDLFFYSERANRDLGAIVYCVGVKDFNETQLARIADSKDHVF 199  
 Db 61 rtasvitaltdgelhedlffysereanrsdlgaivycvgvkdfnetqlariadskdhvf 120

Qy 200 PVNDGFOALQGIHHSILKSCIEILAAEPSPTICAGESQVVRGNGPFHARNVDRVLCSF 259  
 Db 121 pvnldgfoalqgihsilkskscieilaaepsticagesiqvvrrngnfharndrvlcsf 180

Qy 260 KINDSVTLNE 269  
 Db 181 kindsvtlslk 190

RESULT 6  
 AAB18456  
 ID AAB18456 standard; Protein: 488 AA.

XX AAB18456;

DT 15-JAN-2001 (first entry)

XX A human TANGO 216 polypeptide clone.

XX TANGO 266; TANGO 216;; TANGO 261; TANGO 262; TANGO 267;  
 KW cellular proliferation; cellular differentiation; cellular adhesion;  
 KW von Willebrand factor-associated disorder; cell trafficking; cancer;  
 KW hematopoietic associated disease; atelektasis; pulmonary congestion;  
 KW oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;  
 KW intestinal disorder; spleen associated disease; renal disorder;  
 KW cardiovascular disorder; ischemic heart disease; hydrocephalus;  
 KW brain herniation; latrogenic disease; inflammation; meningitis;  
 KW Alzheimer's disease; cerebral toxoplasmosis; Parkinson's disease;  
 KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.

OS Homo sapiens.

XX WO200052022-A1.

XX 08-SEP-2000.

PD

XX 01-MAR-2000; 2000WO-US05226.  
 PF  
 XX 01-MAR-1999; 99US-0122458.  
 PR  
 XX (MILL-) MILLENNIUM PHARM INC.  
 PA  
 XX Barnes TM, Holtzman DA, Sharp JD, Fraser CC;  
 PI  
 XX WPI; 2000-579269/54.  
 DR  
 DR N-PSDB; AAA75158.  
 XX  
 PT Novel human and murine secreted proteins designated TANGO 216, 261,  
 PT 262, 266 and 267 useful as modulating agents of cellular processes,  
 PT e.g. for treating cancer -  
 PT  
 XX Disclosure; Page -: 175pp; English.  
 PS  
 CC AAB18455-57 represent human TANGO 216 proteins. The specification also  
 CC describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO  
 CC polypeptides can be used to modulate cellular proliferation, modulate  
 CC cellular differentiation and/or modulate cellular adhesion. The  
 CC proteins can be used to treat any von Willebrand factor-associated  
 CC disorder, regulate extracellular matrix structuring, cellular adhesion,  
 CC and cell trafficking and/or migration, modulate cellular interactions,  
 CC modulate cell adhesion in proliferative disorders, such as cancer,  
 CC modulate the proliferation, differentiation, and/or function of cells  
 CC that appear in the bone marrow, and leukocytes, treat bone marrow, blood  
 CC and hematopoietic associated diseases and disorders, atelektasis,  
 CC pulmonary congestion or oedema, emphysema, chronic bronchitis, bronchial  
 CC asthma and bronchiectasis, intestinal disorders, spleen associated  
 CC diseases, modulate renal disorders, treat cardiovascular disorders such  
 CC as ischemic heart disease, modulate the proliferation, differentiation,  
 CC and/or function of bone and cartilage cells and to treat bone and/or  
 CC cartilage associated diseases or disorder. They may also be used to  
 CC treat disorders associated with the ovaries, and cerebral oedema,  
 CC hydrocephalus, brain herniations, latrogenic disease, inflammations,  
 CC bacterial and viral meningitis, Alzheimer's Disease, cerebral  
 CC toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers,  
 CC hydrocephalus and encephalitis, and treat hepatic disorders.  
 CC note: the present sequence does not appear in the specification; it was  
 CC created using information provided.  
 CC  
 XX Sequence 488 AA;

Query Match 50.1%; Score 958.5; DB 21; Length 488;  
 Best Local Similarity 51.6%; Pred. No. 2.1e-93;  
 Matches 189; Conservative 63; Mismatches 107; Indels 7; Gaps 3;

Qy 4 AERRALGIGFOWLSLATVILICAGGG--RREDGGPACVGGFDLYFLDKSGSVLHHNNEI 62  
 Db 3 aersparspgswlfpglwllvispggllraqdqpsccrafdlyfvldksvsvannwlei 62

Qy 63 YFVFEQLAHKFIQPLRMSFTVSTRGTTMLKLTEDRQIRQGLEELQKVLPGDVTYME 122  
 Db 63 ynfvqqlaerfvpemrlsfivfssqatillptdgrgkiskgledlkrvpvgetylhe 122

Qy 123 GFERASQIYYENRQGYRTASVITALTGDELHEDLFFYSERANRDLGAIVYCVGVK 182  
 Db 123 glkltaneql--qkagglktssilialtdgklldgfvpsyaekaisrslgasvycvgvld 180

Qy 183 FNEQTARIADSKDHVPVNDGFOALQGIHHSILKSCIEILAAEPSPTICAGESQVVR 242  
 Db 181 feqqlerladskeqvpvkqgffgalkglnsiladsctelleiqpsvsvcvegfqlvis 240

Qy 243 GNGFRHARNVDRVLCSEKINDSVTLNEKPFVSVDYTLCPAPILKEVGMKAALQVSMNDG 302  
 Db 241 grgfmlgsrngsvlctytnetytsvkvpsvqnsmlcpapilnkagetldvsvsfng 300

Qy 303 LSFISSSVITTTCHCDGSIILAIALLILFLALLALMLWFPLCCTVIKEVPPP-----P 358  
 Db 301 ksvisgslivtatecngiaaillvllvlllllgllmwfplcckvkvikpppppppp 360

QY	359	AESEEE 364	factor-associated disorder, regulate extracellular matrix structuring, cellular adhesion, and cell trafficking and/or migration, modulate cellular interactions, modulate cell adhesion in proliferative disorders, such as cancer, modulate the proliferation, differentiation, and/or function of cells that appear in the bone marrow, and leukocytes, treat bone marrow, blood and hematopoietic associated diseases and disorders, atelectasis, pulmonary congestion or oedema, emphysema, chronic bronchitis, bronchial asthma and bronchiectasis, intestinal disorders, spleen associated diseases, modulate renal disorders, treat cardiovascular disorders such as ischemic heart disease, modulate the proliferation, differentiation, and/or function of bone and cartilage cells and to treat bone and/or cartilage associated diseases or disorder. They may also be used to treat disorders associated with the ovaries, cerebral oedema, hydrocephalus, brain herniations, iatrogenic disease, inflammations, bacterial and viral meningitis, Alzheimer's disease, cerebral toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers, hydrocephalus and encephalitis, and treat hepatic disorders.
XX	Sequence	488 AA;	
QY	4	AERRALGIGFQWLSLTLVLCAGGG-RREDGGPACYGFDLYFLDKSGSVLHHWNEI 62	
DB	3	aerspaspqswlfpqglwlvvisgpgllraqeqpcrrafdyfldksgsvannwiei 62	
QY	63	YFVEQLAHKFTSPQLRMSFIVFSTRGTTLMKLTEDREIQRLGELQKVLPGGDTYMH 122	
DB	63	ynfvqqlaerfvspemlsfivfssqatiilptdrgkiskgledkrsvpvgetyhe 122	
QY	123	GFRASEQIYYENRQYRTASVIIALTDELHEDLFFYSEREANRSRLGAIYVCVGVKD 182	
DB	123	glkianeqi--qkagglktssilialtdgklglvpsyaekaklsrsigasyvcgvld 180	
QY	183	FNETQLARIADSKDHVPVNDGFOALQGIITHSLKKSCIEILAAEPSTICAGESFOVV 242	
DB	181	fedaqleriadskqvfpvkgfgalkgiinslaqscleilqlpsvscvgeefdivis 240	
QY	243	GNCFRHNVDVRLCSFKINDSVTLNEKPFSEVEDTYLLCPAPILKEVKMKAALQVSMNDG 302	
DB	241	grgfmlgsrnsgsvlctytnetyttsvkpsvqslmclpcaplinkagetldsvsfng 300	
QY	303	LSFTSSSVIITTHCSDGSILATLILFLLLALILMFWPCLCCTVITKEVPPP---P 358	
DB	301	kvisgsliivtatecsngiaaivilvlllllgllmwfplckkvikdppppppap 360	
QY	359	AESEEE 364	
DB	361	keeeee 366	
RESULT	8		
AAAB18455			
ID	AAAB18455	standard; Protein; 488 AA.	
XX	AAAB18455;		
AC	AAAB18455;		
XX	15-JAN-2001	(first entry)	
DT	15-JAN-2001	(first entry)	
DE	A human	TANGO 216 polypeptide clone.	
XX	TANGO 266;	TANGO 216;; TANGO 261; TANGO 262; TANGO 267;	
KW	cellular proliferation; cellular differentiation; cellular adhesion;		
KW	von Willebrand factor-associated disorder; cell trafficking; cancer;		
KW	hematopoietic associated disease; atelectasis; pulmonary congestion;		
KW	hematopoietic associated disease; atelectasis; bronchial asthma; bronchiectasis;		
KW	oedema; emphysema; chronic bronchitis; atelectasis; pulmonary congestion;		
KW	intestinal disorder; spleen associated disease; renal disorder;		
KW	cardiovascular disorder; ischemic heart disease; hydrocephalus;		
KW	brain herniation; iatrogenic disease; inflammation; meningitis;		
KW	Alzheimer's disease; cerebral toxoplasmosis; Parkinson's disease;		
KW	multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.		
XX	Homo sapiens.		
XX	Location/Qualifiers		
PH	Key	1..34	
FT	Peptide	/note= "signal peptide"	
FT	Domain	34..79	
FT	Domain	/note= "extracellular domain"	
FT	Protein	35..488	
FT	Domain	/note= "mature protein"	
FT	Domain	44..213	
FT	Domain	/note= "von Willebrand factor A domain"	
FT	Domain	80..97	
FT	Domain	/note= "transmembrane domain"	
FT	Domain	98..317	
FT	Domain	/note= "cytoplasmic domain"	
FT	Domain	318..341	
FT	Domain	/note= "transmembrane domain"	
FT	Domain	342..488	
FT	Domain	/note= "extracellular domain"	
XX	WO200052022-A1.		
XX	08-SEP-2000.		
XX	01-MAR-2000;	2000WO-US05226.	
XX	01-MAR-1999;	99US-0122458.	
XX	(MILL-) MILLENNIUM PHARM INC.		
XX	Barnes TM, Holtzman DA, Sharp JD, Fraser CC;		
XX	WPI: 2000-579269/54.		
DR	N-PSDB; AAA75149.		
XX	Novel human and murine secreted proteins designated TANGO 216, 261,		
PT	262, 266 and 267 useful as modulating agents of cellular processes,		
PT	e.g. for treating cancer -		
XX	Claim 8; Fig 1A-C; 175pp; English.		
XX	The present sequence represents a human TANGO 216 polypeptide. The		
CC	specification also describes TANGO 266, TANGO 261, TANGO 262, and		
CC	TANGO 267. The TANGO polypeptides can be used to modulate cellular		
CC	proliferation, modulate cellular differentiation and/or modulate		
CC	cellular adhesion. The proteins can be used to treat any von Willebrand		



Fri Aug 9 10:56:58 2002

us-09-970-076-2.1ag

```
CC bacterial and viral meningitis, Alzheimer's Disease, cerebral
CC toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers,
CC hydrocephalus and encephalitis, and treat hepatic disorders.
CC note: the present sequence does not appear in the specification; it was
CC created using information provided.
XX
SQ Sequence 488 AA;

Query Match 49.9%; Score 954.5; DB 21; Length 488;
Best Local Similarity 51.4%; Pred. No. 5,6e-93;
Matches 188; Conservative 64; Mismatches 107; Indels 7; Gaps 3;

QY 4 AERRALGICFQWLSLATVLCAGOGG-RREDGGPACYGDFLYFLDLSGSLVHHWNEI 62
Db 3 aersparspgswlfpgllwllvsgpgllraqeqscrrafldlyfldksgvaunwidi 62
QY 63 YFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQIROGLELQRLVPGDITYMHE 122
Db 63 ynfvqqlaerfivspemrisfivssgatilpltdgkiskgledkkrvspgetyibe 122
QY 123 GFERASQIYYENROGYRTASVIAITDGLHEDLFYERERANRSDILGAIVYCVGVKD 182
Db 123 glklaneqi--qkagglktssilialtdgklglvpsyaekesirsigasvycvgvid 180
QY 183 FNETQLARIADSKDHPVNDGFCALQGIITHSLKSCIEILAAEPSTICAGESFQWVVR 242
Db 181 feqagleriadskqvfpvkggfgalklnslldgcteilqlpssvcvgeefivls 240
QY 243 GNGFRHARNVDRLCSFKINDSVTLNKPFSVEDTYLLCPAPILKEVGMKAALQVSMNDG 302
Db 241 grgfmgsrsgsvlctytnetytttsvkpsvqinsmlcpapilnkagetldvsvfng 300
QY 303 LSFTSSSVITTHCDSGSIATALLIFILLALALWFEWPLCCTVIIKEVPPP---P 358
Db 301 ksvigsliivtatscngiaailvlllllglgmwifwplcckvvikdpppppppap 360
QY 359 AESEEE 364
Db 361 keeeee 366

RESULT 10
AAU19662
ID AAU19662 standard; Protein; 587 AA.
XX
AC AAU19662;
XX
DT 06-DEC-2001 (first entry)
XX
DE Human novel extracellular matrix protein, Seq ID No 312.
XX
KW Human; secreted extracellular matrix protein; immunomodulatory;
KW Anti-HIV; antianemic; antirheumatic; antisclerotic; cardiac; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytostatic;
KW antialzheimers; immune/autoimmune disease; HIV infection; anaemia;
KW human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;
KW cancers; hyperproliferative disorder; breast neoplasm; melanoma;
KW Sezary syndrome; Gaucher's disease; neurological diseases;
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
KW cardiac arrest; tachycardia; angina; infection; corneal infections;
KW wound healing; immunogen; gene therapy; antisense; food additive.
XX
OS Homo sapiens.
XX
PN WO200155368-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01348.
XX
PP 31-JAN-2000; 2000US-0179065.
XX
PR 04-FEB-2000; 2000US-0180628.
PR
```

```
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR
```



KW Alzheimer's Disease; cerebral toxoplasmosis; Parkinson's disease;  
KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.  
XX  
OS Mus sp.  
XX  
PN WO200052022-A1.  
XX  
XX  
PD 08-SEP-2000.  
XX  
PF 01-MAR-2000; 2000WO-US05226.  
XX  
XX 01-MAR-1999; 99US-0122458.  
PR  
XX (MILL-) MILLENNIUM PHARM INC.  
PA  
XX Barnes TM, Holtzman DA, Sharp JD, Fraser CC;  
PI  
XX WPI: 2000-579269/54.  
XX N-PSDB; AAA75160.  
DR  
XX Novel human and murine secreted proteins designated TANGO 216, 261,  
PT 262, 266 and 267 useful as modulating agents of cellular processes,  
PT e.g. for treating cancer -  
XX  
PS Disclosure; Page -: 175pp; English.  
XX  
XX AAB18458-60 represent murine TANGO 216 proteins. The specification also  
CC describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO  
CC polypeptides can be used to modulate cellular proliferation, modulate  
CC cellular differentiation and/or modulate cellular adhesion. The  
CC proteins can be used to treat any von Willebrand factor-associated  
CC disorder, regulate extracellular matrix structuring, cellular adhesion,  
CC and cell trafficking and/or migration, modulate cellular interactions,  
CC modulate cell adhesion in proliferative disorders, such as cancer,  
CC modulate the proliferation, differentiation, and/or function of cells  
CC that appear in the bone marrow, and leukocytes, treat bone marrow, blood  
CC and hematopoietic associated diseases and disorders, atelectasis,  
CC pulmonary congestion or oedema, emphysema, chronic bronchitis, bronchial  
CC asthma and bronchiectasis, intestinal disorders, spleen associated  
CC diseases, modulate renal disorders, treat cardiovascular disorders such  
CC as ischemic heart disease, modulate the proliferation, differentiation,  
CC and/or function of bone and cartilage cells and to treat bone and/or  
CC cartilage associated diseases or disorders. They may also be used to  
CC treat disorders associated with the ovaries, and cerebral oedema,  
CC hydrocephalus, brain herniations, iatrogenic disease, inflammations,  
CC bacterial and viral meningitis, Alzheimer's Disease, cerebral  
CC toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers,  
CC hydrocephalus and encephalitis, and treat hepatic disorders.  
CC note: the present sequence does not appear in the specification; it was  
CC created using information provided.  
XX  
SQ Sequence 487 AA;

Query Match 49.0%; Score 937.5; DB 21; Length 487;  
Best Local Similarity 50.1%; Pred. No. 3.7e-91;  
Matches 186; Conservative 63; Mismatches 115; Indels 7; Gaps 4;

QY 1 MATAERRALGIGFOWLSLATVILICAGOGG-RREDGPGACYGGFDLYFLDKSGSVLHHW 59  
Db 1 mvagrsrarspg-swlfpgllwllavggpgsllqadqpsckkafdlyfldksgsvannw 59  
QY 60 NEIYFVQLAHKFTSPOLRMSFVFTRGTTMLKLTEDRQIROGLEELQKVLPGDITY 119  
Db 60 ieiynfvhqlterfvspeirnsifvssqatliilptgdrykigkglcdlkavkpvgety 119  
QY 120 MHEGFERASEQIYYENRGYRTASVIIATDCEHLEDLFFYSERANRSRLGAIVCVG 179  
Db 120 ihegxlkaneql--qnaqgklkassiiatldqkldglwpsyaenakksrslgasvycv 177  
QY 180 VKDFNETQIARIADSKDHFVPVNDGFQALQGTIHSILKSCITEILAAEPSTICAGESFQV 239  
Db 178 vldfeaqelriadskdgvfkggfgalkginsilaqsciteilelpsvssvcgkfgv 237

QY 240 VVRGNGFHARNVDVRLGCSFKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSM 299  
Db 238 vltgravtsishdgsvlctftanstytksekpvsilcpapvlnkdgetlevsisy 297  
QY 300 NDGLSFISSTVIITTHCSGSLAIAALLILFLALLALMMFWPLCCTVTIIEKVPVPPPA 359  
Db 298 ndgksavsrsltitactectngiaaivaillvlllllgaalmmwfwplcckvvikdppppps 357  
QY 360 ---ESENKI 367  
Db 358 apmeeeedpi 368  
RESULT 12  
AAB18448  
ID AAB18448 standard; Protein; 487 AA.  
XX  
AC AAB18448;  
XX  
DT 15-JAN-2001 (first entry)  
XX  
DE Amino acid sequence of a murine TANGO 216 polypeptide.  
XX  
KW TANGO 266; TANGO 216;; TANGO 261; TANGO 262; TANGO 267;  
KW cellular proliferation; cellular differentiation; cellular adhesion;  
KW von Willebrand factor-associated disorder; cell trafficking; cancer;  
KW hematopoietic associated disease; atelectasis; pulmonary congestion;  
KW oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;  
KW intestinal disorder; spleen associated disease; renal disorder;  
KW cardiovascular disorder; ischemic heart disease; hydrocephalus;  
KW brain herniation; iatrogenic disease; inflammation; meningitis;  
KW Alzheimer's Disease; cerebral toxoplasmosis; Parkinson's disease;  
KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.  
XX  
Mus sp.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..33 "signal peptide"  
FT Domain 34..79 "extracellular domain"  
FT Protein 34..487 "mature protein"  
FT Domain 80..97 "transmembrane domain"  
FT Domain 98..317 "transmembrane domain"  
FT Domain 318..341 "cytoplasmic domain"  
FT Domain 342..487 "transmembrane domain"  
FT Domain /note= "extracellular domain"  
XX  
PN WO200052022-A1.  
XX  
XX 08-SEP-2000.  
XX  
PF 01-MAR-2000; 2000WO-US05226.  
XX  
XX 01-MAR-1999; 99US-0122458.  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
XX  
XX Barnes TM, Holtzman DA, Sharp JD, Fraser CC;  
PI  
XX WPI: 2000-579269/54.  
XX N-PSDB; AAA75150.  
DR  
XX Novel human and murine secreted proteins designated TANGO 216, 261,  
PT 262, 266 and 267 useful as modulating agents of cellular processes,  
PT e.g. for treating cancer -  
XX  
XX Claim 8; Fig 2A-C; 175pp; English.  
PS

XX The present sequence represents a murine TANGO 216 polypeptide. The  
 CC specification also describes TANGO 266, TANGO 261, TANGO 262, and  
 CC TANGO 267. The TANGO polypeptides can be used to modulate cellular  
 CC proliferation, modulate cellular differentiation and/or modulate  
 CC cellular adhesion. The proteins can be used to treat any von Willebrand  
 CC factor-associated disorder, regulate extracellular matrix structuring,  
 CC cellular adhesion, and cell trafficking and/or migration, modulate  
 CC cellular interactions, modulate cell adhesion in proliferative  
 CC disorders, such as cancer, modulate the proliferation, differentiation,  
 CC and/or function of cells that appear in the bone marrow, and leukocytes,  
 CC treat bone marrow, blood and hematopoietic associated diseases and  
 CC disorders, atelectasis, pulmonary congestion or oedema, emphysema,  
 CC chronic bronchitis, bronchial asthma and bronchiectasis, intestinal  
 CC disorders, spleen associated diseases, modulate renal disorders, treat  
 CC cardiovascular disorders such as ischemic heart disease, modulate the  
 CC proliferation, differentiation, and/or function of bone and cartilage  
 CC cells and to treat bone and/or cartilage associated diseases or  
 CC disorder. They may also be used to treat disorders associated with the  
 CC ovaries, cerebral oedema, hydrocephalus, brain herniations, iatrogenic  
 CC disease, inflammations, bacterial and viral meningitis, Alzheimer's  
 CC disease, cerebral toxoplasmosis, Parkinson's disease, multiple sclerosis,  
 CC brain cancers, hydrocephalus and encephalitis, and treat hepatic  
 CC disorders.  
 XX  
 SQ Sequence 487 AA;

Query Match 48.9%; Score 936.5; DB 21; Length 487;  
 Best Local Similarity 50.1%; Pred. No. 4.7e-91;  
 Matches 186; Conservative 63; Mismatches 115; Indels 7; Gaps 4;

QY 1 MATAERRALGIGFQWLSLATLVLCAGQGG-RREDGGPACYGGEFDLYFTLDKSGSVLHHW 59  
 DB 1 mvagrsrarspg-swlfpglwlavggpslllgaqeqpsckafdlfvlidksqsvannw 59

QY 60 NEIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLELOKVLPGDPT 119  
 DB 60 ielynfvhqltrfsvpmrlsfivfssqatlilptgdrykigkqledkavkpvgety 119

QY 120 MHEGFERASEQIYYENROGYRTASVITALTDCGELHEDLFFYSERANRSRDGAIVYCVG 179  
 DB 120 ihegiklaneql-qnaggikassilialtdgldglvpsaeneakksrsgasvycvg 177

QY 180 VKDFNETQLARIADSKDHVPVNDGFOALQGIHSTLKSKSTEILAAEPSTICAGESFOV 239  
 DB 178 vldfeaqqlerladskdgvfpkvgfalkginsllaqscsteilelspssvcvgekfv 237

QY 240 VVRNGFRHARNVDVRLCSFKINDSVTLNEKPSFVEDTYLLCPAPILKEVGMKAALQVSM 299  
 DB 238 vltgravtsishdgsvlctftanstytksekpsvsiqpsilcpapvlnkdgetlevisiy 297

QY 300 NDLSPFISSSVIITTHCSGSLATALLLLELLALALLWFWPLCCVVIKEVPPPPA 359  
 DB 298 ndgksavrsrltitactcnglaaivailvlllilglaamwfwplcckvkvdkppppps 357

QY 360 ---EESSENKI 367  
 DB 358 apmeeeedpl 368

RESULT 13  
 AAB18460  
 ID AAB18460 standard; Protein; 487 AA.  
 XX  
 AC AAB18460;  
 XX  
 DT 15-JAN-2001 (first entry)  
 XX  
 DE A murine TANGO 216 polypeptide clone.  
 XX  
 KW TANGO 266; TANGO 216; TANGO 261; TANGO 262; TANGO 267;  
 KW cellular proliferation; cellular differentiation; cellular adhesion;

KW von Willebrand factor-associated disorder; cell trafficking; cancer;  
 KW hematopoietic associated disease; atelectasis; pulmonary congestion;  
 KW oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;  
 KW intestinal disorder; spleen associated disease; renal disorder;  
 KW cardiovascular disorder; ischemic heart disease; hydrocephalus;  
 KW brain herniation; iatrogenic disease; inflammation; meningitis;  
 KW Alzheimer's disease; cerebral toxoplasmosis; Parkinson's disease;  
 KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.  
 XX  
 OS Mus sp.  
 XX  
 PN WO200052022-A1.  
 XX  
 PD 08-SEP-2000.  
 XX  
 PF 01-MAR-2000; 2000WO-US05226.  
 XX  
 PR 01-MAR-1999; 99US-0122458.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Barnes TM, Holtzman DA, Sharp JD, Fraser CC;  
 XX  
 DR WPI; 2000-579269/54.  
 DR N-PSDB; AAA75162.  
 XX  
 PT Novel human and murine secreted proteins designated TANGO 216, 261,  
 PT 262, 266 and 267 useful as modulating agents of cellular processes,  
 PT e.g. for treating cancer -  
 XX  
 PS Disclosure: Page -: 175pp; English.

CC AAB18458-60 represent murine TANGO 216 proteins. The specification also  
 CC describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO  
 CC polypeptides can be used to modulate cellular proliferation, modulate  
 CC cellular differentiation and/or modulate cellular adhesion. The  
 CC proteins can be used to treat any von Willebrand factor-associated  
 CC disorder, regulate extracellular matrix structuring, cellular adhesion,  
 CC and cell trafficking and/or migration, modulate cellular interactions,  
 CC modulate cell adhesion in proliferative disorders, such as cancer,  
 CC modulate the proliferation, differentiation, and/or function of cells  
 CC that appear in the bone marrow, and leukocytes, treat bone marrow, blood  
 CC and hematopoietic associated diseases and disorders, atelectasis,  
 CC pulmonary congestion or oedema, emphysema, chronic bronchitis, bronchial  
 CC asthma and bronchiectasis, intestinal disorders, spleen associated  
 CC diseases, modulate renal disorders, treat cardiovascular disorders such  
 CC as ischemic heart disease, modulate the proliferation, differentiation,  
 CC and/or function of bone and cartilage cells and to treat bone and/or  
 CC cartilage associated diseases or disorder. They may also be used to  
 CC treat disorders associated with the ovaries, and cerebral oedema,  
 CC hydrocephalus, brain herniations, iatrogenic disease, inflammations,  
 CC bacterial and viral meningitis, Alzheimer's Disease, cerebral  
 CC toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers,  
 CC hydrocephalus and encephalitis, and treat hepatic disorders.  
 CC note: the present sequence does not appear in the specification; it was  
 CC created using information provided.  
 XX  
 SQ Sequence 487 AA;

Query Match 48.9%; Score 935.5; DB 21; Length 487;  
 Best Local Similarity 50.1%; Pred. No. 6e-91;  
 Matches 186; Conservative 63; Mismatches 115; Indels 7; Gaps 4;

QY 1 MATAERRALGIGFQWLSLATLVLCAGQGG-RREDGGPACYGGEFDLYFTLDKSGSVLHHW 59  
 DB 1 mvagrsrarspg-swlfpglwlavggpslllgaqeqpsckafdlfvlidksqsvannw 59

QY 60 NEIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLELOKVLPGDPT 119  
 DB 60 ielynfvhqltrfsvpmrlsfivfssqatlilptgdrykigkqledkavkpvgety 119

QY 120 MHEGFERASEQIYYENROGYRTASVITALTDCGELHEDLFFYSERANRSRDGAIVYCVG 179

Db 120 ihelklaneqi--qagglkassilialtdgklidgvlpsyaenacksrslgasvycv 177  
 Qy 180 VKDFNETQLARIADSKDHVPVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQV 239  
 Db 178 vldfeqaqler-iadskdqvpkggfgalkgilnsilaqscetlelspssvvcgekfv 237  
 Qy 240 VVRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSEDTYLLCPAPILKEVGMKAALQVSM 299  
 Db 238 vltgravtsishdgsvlctftanstytksekvslqpsilcpavlnkdgetlevsly 297  
 Qy 300 NDGLSFSSSVIIITTHCSGSIILAIALLILFLALLLWFWPLCCTVIKEVPPPPA 359  
 Db 298 ndgksavrsititactectngiaaivailvlllllgaalmwfwplcckvvikppppps 357  
 Qy 360 ---EESSENKI 367  
 Db 358 apmeeeedpl 368  
 RESULT 14  
 AAB18459  
 ID AAB18459 standard; Protein; 487 AA.  
 AC AAB18459;  
 XX 15-JAN-2001 (first entry)  
 DT A murine TANGO 216 polypeptide clone.  
 DE TANGO 266; TANGO 216; TANGO 262; TANGO 267;  
 XX cellular proliferation; cellular differentiation; cellular adhesion;  
 KW von Willebrand factor-associated disorder; cell trafficking; cancer;  
 KW hematopoietic associated disease, atelectasis; pulmonary congestion;  
 KW oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;  
 KW intestinal disorder; spleen associated disease; renal disorder;  
 KW cardiovascular disorder; ischemic heart disease; hydrocephalus;  
 KW brain herniation; iatrogenic disease; inflammation; meningitis;  
 KW Alzheimer's disease; cerebral toxoplasmosis; Parkinson's disease;  
 KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.  
 OS Mus sp.  
 XX WO200052022-A1.  
 PN 08-SEP-2000.  
 PD 01-MAR-2000; 2000WO-US05226.  
 XX 01-MAR-1999; 99US-0122458.  
 XX (MILL-) MILLENNIUM PHARM INC.  
 XX Barnes TM, Holtzman DA, Sharp JD, Fraser CC;  
 PI WPI; 2000-579269/54.  
 XX N-PSDB; AAA75161.  
 DR Novel human and murine secreted proteins designated TANGO 216, 261,  
 PT 262, 266 and 267 useful as modulating agents of cellular processes,  
 PT e.g. for treating cancer -  
 XX Disclosure; Page -: 175pp; English.  
 PS AAB18458-60 represent murine TANGO 216 proteins. The specification also  
 CC describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO  
 CC polypeptides can be used to modulate cellular proliferation, modulate  
 CC cellular differentiation and/or modulate cellular adhesion. The  
 CC proteins can be used to treat any von Willebrand factor-associated  
 CC disorder, regulate extracellular matrix structuring, cellular adhesion,  
 CC and cell trafficking and/or migration, modulate cellular interactions,  
 CC modulate cell adhesion in proliferative disorders, such as cancer,  
 CC modulate the proliferation, differentiation, and/or function of cells

CC that appear in the bone marrow, and leukocytes, treat bone marrow, blood  
 CC and hematopoietic associated diseases and disorders, atelectasis,  
 CC pulmonary congestion or oedema, emphysema, chronic bronchitis, bronchial  
 CC asthma and bronchiectasis, intestinal disorders, spleen associated  
 CC diseases, modulate renal disorders, treat cardiovascular disorders such  
 CC as ischemic heart disease, modulate the proliferation, differentiation,  
 CC and/or function of bone and cartilage cells and to treat bone and/or  
 CC cartilage associated diseases or disorder. They may also be used to  
 CC treat disorders associated with the ovaries, and cerebral oedema,  
 CC hydrocephalus, brain herniations, iatrogenic disease, inflammations,  
 CC bacterial and viral meningitis, Alzheimer's disease, cerebral  
 CC toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers,  
 CC hydrocephalus, and encephalitis, and treat hepatic disorders.  
 CC note: the present sequence does not appear in the specification; it was  
 CC created using information provided.  
 XX Sequence 487 AA;  
 SQ  
 Query Match 48.8%; Score 933.5; DB 21; Length 487;  
 Best Local Similarity 49.9%; Pred. No. 9,9e-91;  
 Matches 185; Conservative 64; Mismatches 115; Indels 7; Gaps 4;  
 Qy 1 MATERRALGIGFOWLSLATLVLICAGGG--RRDGGPACYGGLYELDKSGVLHHW 59  
 Db 1 mvagrstrarspg-swllfpgllwllavggpgsllqagqpsckakafolyvlcksgsvannw 59  
 Qy 60 NEIYFVEOLAHKFTSPOLRMSFTVFRGTTLMLKLTEDREQIRQGLBELOKVLPGGDTY 119  
 Db 60 idlynfvhlterfvspenrlsfivssqatllpttdrykigkglcdkavkpvgety 119  
 Qy 120 MHEGFERASEQIYYENROGYRTASVIIATDGEHLHEDLFFYSERENRSDLGAIYCVG 179  
 Db 120 ihelklaneqi--qnaaglkassilialtdgklidgvlpsyaenacksrslgasvycv 177  
 Qy 180 VKDFNETQLARIADSKDHVPVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQV 239  
 Db 178 vldfeqaqler-iadskdqvpkggfgalkgilnsilaqscetlelspssvvcgekfv 237  
 Qy 240 VVRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSEDTYLLCPAPILKEVGMKAALQVSM 299  
 Db 238 vltgravtsishdgsvlctftanstytksekvslqpsilcpavlnkdgetlevsly 297  
 Qy 300 NDGLSFSSSVIIITTHCSGSIILAIALLILFLALLLWFWPLCCTVIKEVPPPPA 359  
 Db 298 ndgksavrsititactectngiaaivailvlllllgaalmwfwplcckvvikppppps 357  
 Qy 360 ---EESSENKI 367  
 Db 358 apmeeeedpl 368  
 RESULT 15  
 AAB01428  
 ID AAB01428 standard; Protein; 381 AA.  
 XX AC AAB01428;  
 XX 20-OCT-2000 (first entry)  
 XX Murine TANGO 197.  
 XX TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;  
 KW graft versus-host diseases; rheumatoid arthritis; psoriasis;  
 KW inflammatory bowel disease; septic shock; ulcerative colitis;  
 KW Crohn's disease; chronic myelogenous leukemia; cancer; liver  
 KW disease; Hodgkin's disease; osteoarthritis; Lyme's disease;  
 KW cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;  
 KW systemic lupus erythematosus; transgenic animal; diagnosis;  
 XX prognosis; prophylactic; therapeutic; mouse.  
 OS Mus musculus.



PN WO200039284-A1.  
XX  
PD 06-JUL-2000.  
XX  
XX 23-DEC-1999; 99WO-US31025.  
PF  
XX 30-DEC-1998; 98US-0223546.  
PR  
XX (MILL-) MILLENNIUM PHARM INC.  
PA  
XX Holtzman DA;  
PI  
XX WPI; 2000-465743/40.  
DR N-PSDB; AAA47479.  
XX  
XX Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,  
PT 224 and 239 polypeptides useful for the treatment of asthma, rheumatoid  
PT arthritis, psoriasis and autoimmune diseases  
XX  
XX Claim 8; Fig 27; 209pp; English.  
PS  
XX Nucleic acids encoding TANGO polypeptides are useful as modulating  
CC agents for regulating cellular processes like asthma, graft  
CC versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory  
CC bowel disease, septic shock, ulcerative colitis, Crohn's disease,  
CC chronic myelogenous leukemia, cancer, liver disease, Hodgkin's  
CC disease, osteoarthritis, Lyme's disease, cachexia and autoimmune  
CC diseases e.g. myasthenia gravis, autoimmune diabetes and systemic  
CC lupus erythematosus. The nucleic acids are also useful for producing  
CC transgenic animals and the TANGO polypeptides themselves. Partial  
CC TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in  
CC forensic biology, for diagnostic assays, prognostic assays,  
CC pharmacogenomics and for monitoring clinical trials. TANGO  
CC polypeptides are suitable for both prophylactic and therapeutic  
CC methods for treating a subject at risk of a disorder or having a  
CC disorder associated with aberrant TANGO expression. A wide range  
CC of cellular disorders can be treated.  
XX  
SQ Sequence 381 AA;

Query Match 47.5%; Score 909; DB 21; Length 381;  
Best Local Similarity 98.3%; Pred. No. 2.8e-88;  
Matches 175; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 187 QLARIADSKDHVPVNDGFOALQGIHSHILKKSCTEILAAEPSTICAGESQVVRGNGF 246  
Db 4 qlariadskdhvfpvndgfalqglhshilkkscieilaaepsticagesfqvvvrgngf 63

Qy 247 RHARNVDRVLCSFKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFI 306  
Db 64 rharnvdrvlcsfkindsvtlinekpfavedtyllcpapilkevgmkaalqvsmndglsfi 123

Qy 307 SSSVITTTTHCSGSIILAIALLIFLLALALLMFWFPLCCTVIKEVPPPPAESEE 364  
Db 124 sssvittthcsdgsilaiallvfillalallwfwplcctviikevppppvessee 181

Search completed: August 9, 2002, 10:32:10  
Job time: 129 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:33:17 ; Search time 31.3 Seconds  
(without alignments)  
455.233 Million cell updates/sec

Title: US-09-970-076-2  
Perfect score: 1914  
Sequence: 1 MATERRALGIGFQWLSLAT.....VVIKEVPPPAESENKIK 368

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1894	99.0	564	1 ATR_HUMAN	Q9h6x2 homo sapien
2	1793	93.7	562	1 ATR_MOUSE	Q9cz52 mus musculus
3	651	34.0	386	1 CMG2_HUMAN	P58335 homo sapien
4	159.5	8.3	1162	1 ITAD_HUMAN	Q13349 homo sapien
5	145.5	7.6	1152	1 ITAM_HUMAN	P11215 homo sapien
6	143.5	7.5	1163	1 ITAX_HUMAN	P20702 homo sapien
7	143	7.5	3124	1 CA1C_CHICK	P13944 gallus gall
8	141	7.4	1153	1 ITAM_MOUSE	P05555 mus musculus
9	139	7.3	3067	1 CA1C_MOUSE	Q60847 mus musculus
10	131	6.8	760	1 CO2_MOUSE	P21180 mus musculus
11	131	6.8	3063	1 CA1C_HUMAN	Q93715 homo sapien
12	130	6.8	1888	1 CA1E_CHICK	P32018 gallus gall
13	128	6.7	3051	1 YNK3_CAEEL	P34576 caenorhabdi
14	120.5	6.3	496	1 CAMA_HUMAN	P21941 homo sapien
15	118	6.2	929	1 CA1C_NOTVTI	Q91145 notophthalm
16	117.5	6.1	452	1 MTN3_CHICK	O42401 gallus gall
17	113.5	5.9	764	1 CFAB_HUMAN	P00751 homo sapien
18	113.5	5.9	956	1 MTN9_HUMAN	O00339 homo sapien
19	111.5	5.8	2944	1 CA17_HUMAN	Q02388 homo sapien
20	110.5	5.8	761	1 CFAB_MOUSE	P04186 mus musculus
21	109.5	5.7	1029	1 CA26_MOUSE	Q02788 mus musculus
22	108.5	5.7	500	1 CAMA_MOUSE	P51942 mus musculus
23	108.5	5.7	624	1 MTN4_MOUSE	O89029 mus musculus
24	108.5	5.7	956	1 MTN2_MOUSE	O08746 mus musculus
25	107.5	5.6	493	1 CAMA_CHICK	P05099 gallus gall
26	107	5.6	1189	1 ITAH_HUMAN	Q9ukx5 homo sapien
27	103	5.4	619	1 MTN4_HUMAN	O95460 homo sapien
28	102	5.3	486	1 MTN3_HUMAN	O15232 homo sapien
29	99.5	5.2	597	1 ROP_DROME	Q07327 drosophila
30	99	5.2	341	1 Y173_BORBU	O51195 borrelia bu
31	99	5.2	559	1 TRAP_PLAFA	P16893 plasmodium
32	99	5.2	1018	1 CA26_HUMAN	P12110 homo sapien
33	98	5.1	3176	1 CA36_HUMAN	P12111 homo sapien

34	97.5	5.1	642	1 DNAK_NEIMB	Q9k0n4 neisseria m
35	97	5.1	285	1 ITAL_CHICK	O90615 gallus gall
36	97	5.1	1151	1 ITAL_HUMAN	P56199 homo sapien
37	97	5.1	3137	1 CA36_CHICK	P15989 gallus gall
38	96	5.0	639	1 CA1C_RABIT	Q28902 oryctolagus
39	95.5	5.0	1179	1 ITAE_HUMAN	P38570 homo sapien
40	95.5	5.0	1180	1 ITAL_RAT	P18614 rattus norv
41	95	5.0	1167	1 ITAE_MOUSE	O60677 mus musculus
42	92.5	4.8	401	1 HUTI_CAUCR	P58079 caulobacter
43	92.5	4.8	550	1 COCH_HUMAN	O43405 homo sapien
44	92.5	4.8	642	1 DNAK_NEIMA	Q91vq9 neisseria m
45	91	4.8	1167	1 ITAG_HUMAN	O75578 homo sapien

ALIGNMENTS

```
RESULT 1
ATR_HUMAN
ID ATR_HUMAN STANDARD; PRT; 564 AA.
AC Q9H6X2; Q9NPV3;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Antrax toxin receptor precursor (Tumor endothelial marker 8).
GN ATR OR TEM8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20407466; PubMed=10947988;
RA St Croix B., Rago C., Velculescu V., Traverso G., Romans K.E.,
RA Montgomery E., Lai A., Riggs G.J., Lengauer C., Vogelstein B.,
RA Kinzler K.W.;
RT "Genes expressed in human tumor endothelium.";
RL Science 289:1197-1202(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX PubMed=11700562;
RA Bradley K.A., Mogridge J., Mourez M., Collier R.J., Young J.A.T.;
RT "Identification of the cellular receptor for anthrax toxin.";
RL Nature 414:225-229(2001).
RN [3]
RP SEQUENCE OF 184-564 FROM N.A. (ISOFORM 1), AND SEQUENCE FROM N.A.
RP (ISOFORM 3).
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Kidney;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Cellular role is not yet known. (PA) of Bacillus
CC -!- SUBUNIT: Binds to the protective antigen (PA) of Bacillus
CC anthracis.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -!- ALTERNATIVE PRODUCTS: 4 isoforms; 1 (shown here), 2, 3 and 4; seem
CC to be produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Highly expressed in tumor endothelial cells
CC but not in normal endothelial cells.
CC -!- DOMAIN: Binding to PA seems to be effected through the VWA domain.
CC -!- SIMILARITY: BELONGS TO THE ATR FAMILY.
CC -!- SIMILARITY: CONTAINS 1 VWA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; AF279145; AAK52094.1; -  
 DR EMBL; AK025429; BAB15128.1; ALT\_INIT.  
 DR EMBL; AK001463; BAA91707.1; ALT\_FRAME.  
 DR EMBL; BC012074; AAH12074.1; -  
 DR MM; 606410; -  
 DR InterPro: IPR002035; vWFA.  
 DR SMART: SM00327; vWA: 1.  
 DR PROSITE; PS0234; vWFA: 1.  
 DR Receptor; Transmembrane; Glycoprotein; Signal; Alternative splicing.  
 KW SIGNAL 1 32  
 FT CHAIN 33 564  
 FT DOMAIN 33 321  
 FT TRANSMEM 322 342  
 FT DOMAIN 343 564  
 FT CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 44 215  
 FT vWFA.  
 FT DOMAIN 360 368  
 FT ASP/GLU-RICH (HIGHLY ACIDIC).  
 FT DOMAIN 506 564  
 FT CARBOHYD 166 166  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 184 184  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 262 262  
 FT EDDD -> NKIK (IN ISOFORM 2).  
 FT VARSPPLIC 365 368  
 FT MISSING (IN ISOFORM 2).  
 FT VARSPPLIC 268 297  
 FT NEKPSVEDYLLCAPILKEVGMKAALQV -> SKSIQSP  
 FT WVSSTGFKGNSHPCLPAPRHT (IN ISOFORM 3).  
 FT MISSING (IN ISOFORM 3).  
 FT VARSPPLIC 298 564  
 FT DGSIIAIALILFL -> LHKIASGTTAACME (IN  
 FT ISOFORM 4).  
 FT VARSPPLIC 334 564  
 FT MISSING (IN ISOFORM 4).  
 FT VARSPPLIC 339 333  
 FT B118A00AD5DF2233 CRC64;  
 FT SEQUENCE 564 AA; 62789 MW; B118A00AD5DF2233 CRC64;

Query Match 99.0%; Score 1894; DB 1; Length 564;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-147;  
 Matches 364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATAERRALGIGFQWLSLATLVLCAGQGGREDGGPACYGFDLYFLDKSGSVLHWN 60  
 DB 1 MATAERRALGIGFQWLSLATLVLCAGQGGREDGGPACYGFDLYFLDKSGSVLHWN 60  
 QY 61 EIYFVEQLAHFISPOLRMSFIVSTRTTLMKLTEDREQIRQGLEELQKVLPGGDTYM 120  
 DB 61 EIYFVEQLAHFISPOLRMSFIVSTRTTLMKLTEDREQIRQGLEELQKVLPGGDTYM 120  
 QY 121 HEGFERASQIYYENRQGYRTASVIALTDGELHEDLFFYSREANRSDLGAIYVCVGV 180  
 DB 121 HEGFERASQIYYENRQGYRTASVIALTDGELHEDLFFYSREANRSDLGAIYVCVGV 180  
 QY 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHSILKSCIELAAEPSTICAGESFQVY 240  
 DB 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHSILKSCIELAAEPSTICAGESFQVY 240  
 QY 241 VRNGFRHARNVDRVLCSPKINDSVTLNKPFSVEDTYLLCAPILKEVGMKAALQVSN 300  
 DB 241 VRNGFRHARNVDRVLCSPKINDSVTLNKPFSVEDTYLLCAPILKEVGMKAALQVSN 300  
 QY 301 DGLSFSTSSVITTHCSOGSILAIALILFLALLALWFWPLCCTVIKEVPPPPAE 360  
 DB 301 DGLSFSTSSVITTHCSOGSILAIALILFLALLALWFWPLCCTVIKEVPPPPAE 360  
 QY 361 ESEE 364  
 DB 361 ESEE 364  
 RESULT 2  
 ATR\_MOUSE  
 ID ATR\_MOUSE PRT; 562 AA.  
 AC Q9CZ52;  
 DT 01-MAR-2002 (Rel. 41, Created)

DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Antrax toxin receptor precursor (tumor endothelial marker 8).  
 GN ATR OR TEM8.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=21443268; PubMed=11559528;  
 RA Carson-Walter E.B., Watkins D.N., Nanda A., Vogelstein B.,  
 RA Kinzler K.W., St Croix B.;  
 RA "Cell surface tumor endothelial markers are conserved in mice and  
 RT humans.";  
 RL Cancer Res. 61:6649-6655(2001).  
 RN [2]  
 RN SEQUENCE OF 88-562 FROM N.A. (ISOFORM 2).  
 RC STRAIN=C57BL/6J; TISSUE=Embryo;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RA "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 CC -!- FUNCTION: Cellular role is not yet known.  
 CC -!- SUBUNIT: Binds to the protective antigen (PA) of Bacillus  
 CC anthracis (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).  
 CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; seem to  
 CC be produced by alternative splicing.  
 CC -!- DOMAIN: Binding to PA seems to be effected through the vWA domain  
 CC (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE ATR FAMILY.  
 CC -!- SIMILARITY: CONTAINS 1 vWFA DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF378762; AAL11999.1; -  
 DR EMBL; AK013005; BAB28591.1; ALT\_INIT.  
 DR MGD; MGI:1919432; Atr.  
 DR InterPro: IPR002035; vWFA.  
 DR PROSITE; PS0234; vWFA: 1.  
 DR Receptor; Transmembrane; Glycoprotein; Signal; Alternative splicing.  
 FT SIGNAL 1 30  
 FT CHAIN 31 562  
 FT DOMAIN 31 319  
 FT EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 320 340  
 FT POTENTIAL.  
 FT DOMAIN 341 562  
 FT CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 42 213  
 FT vWFA.  
 FT DOMAIN 358 366  
 FT ASP/GLU-RICH (HIGHLY ACIDIC).  
 FT DOMAIN 501 562  
 FT PRO-RICH.

```
FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 477 562 GCINTRVKNQSPAKYPLNNTYHSPSPAPITPTPPAP
FT FT HCPPAPASPTPTPPPPSTLPPPPQAPPNRAPPSPRPP
FT FT RPSV -> RFRGWLITICLSKHVHFRGDKGPETPLKOA
FT FT WMFSSFLERAFQ (IN ISOFORM 2).
FT FT
SQ SEQUENCE 562 AA; 62308 MW; 6AC92049B4BB4F7C CRC64;

Query Match 93.7%; Score 1793; DB 1; Length 562;
Best Local Similarity 96.1%; Pred. No. 5.6e-139;
Matches 342; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 9 LGIGFOWLSLTLVLCACAGGREDGPGACYGDFLYFLDKSGSVLHHWNEIYFVEQ 68
DB 7 LGAGLGLCVAAVLVLCAGHGRREDGPGACYGDFLYFLDKSGSVLHHWNEIYFVEQ 66
QY 69 LAHFISPOLRMSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFERAS 128
DB 67 LAHFISPOLRMSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFERAS 126
QY 129 EQIYYENRQGYRTASVIATLDGELHEDLFFYSREANRSDGAIYCVGVKDFNETQL 188
DB 127 EQIYYENRQGYRTASVIATLDGELHEDLFFYSREANRSDGAIYCVGVKDFNETQL 186
QY 189 ARIADSKDHVPVNDGFOALQGIHSLKSCIEILAAEPSTICAGSFQVVGNGFRH 248
DB 187 ARIADSKDHVPVNDGFOALQGIHSLKSCIEILAAEPSTICAGSFQVVGNGFRH 246
QY 249 ARNVDRLVCLSPKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQVSNNDGLSFSS 308
DB 247 ARNVDRLVCLSPKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQVSNNDGLSFSS 306
QY 309 SVITTHCHSDGSLATALLFLLLALLLWFWPLCCTVIKEVPPPPAESEEE 364
DB 307 SVITTHCHSDGSLATALLFLLLALLLWFWPLCCTVIKEVPPPPAESEEE 362

RESULT 3
CMG2_HUMAN
ID CMG2_HUMAN STANDARD; PRT; 386 AA.
AC P58335;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Capillary morphogenesis protein-2 precursor (CMG-2).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21539596; PubMed=11693410;
RA Bell S.E., Mavila A., Salazar R., Bayless K.J., Kanagala S.,
RA Maxwell S.A., Davis G.E.;
RA "Differential gene expression during capillary morphogenesis in 3D
RT collagen matrices: regulated expression of genes involved in basement
RT membrane matrix assembly, cell cycle progression, cellular
RT differentiation and G-protein signaling.";
RL J. Cell Sci. 114:2755-2773(2001).
CC -1- SUBUNIT: Seems to bind to collagen type IV and laminin.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE ATR FAMILY.
CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL: AY040326; AAK77222.1; .
DR PROSITE: PS50234; VWFA; 1.
KW Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 386 CAPILLARY MORPHOGENESIS PROTEIN-2.
FT DOMAIN 34 215 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 216 236 POTENTIAL.
FT DOMAIN 237 386 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 44 213 VWFA.
FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 386 AA; 42918 MW; 0F9B2D6688EAB17A CRC64;

Query Match 34.0%; Score 651; DB 1; Length 386;
Best Local Similarity 38.8%; Pred. No. 7.3e-46;
Matches 142; Conservative 45; Mismatches 69; Indels 110; Gaps 4;

QY 4 AERRALGIGFOWLSLTLVLCACAGG-RREDGGPGACYGDFLYFLDKSGSVLHHWNEI 62
DB 3 AERSPARSPGSLPGLWLLVLSGPGGLLRAEQEPSCRRAFDLYFLDKSGSVANNMIEI 62
QY 63 YFVFEQLAHFISPOLRMSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDGYMH 122
DB 63 YNFVQQLAERFVSPERMSLFIYFSSQATIIILPTCDRGKISKGLDLKRVSPVGETYI 122
QY 123 GFERASEQIYYENRQGYRTASVIATLDGELHEDLFFYSREANRSDGAIYCVGVK 182
DB 123 GLKLANEOI--QKAGGLKTSIIIALTDGKLDGLVPSYAEKEAKISGLGASVYCVGVLD 180
QY 183 FNETQLARIADSKDHVPVNDGFOALQGIHSLKSCIEILAAEPSTICAGSFQVVG 242
DB 181 FEQQLERADSKDQVFPVKGFOALGINS----- 212
QY 243 GNGFRHARNVDRLVCLSPKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQV 302
DB 213 ----- 212
QY 303 LSFISSTVITTHCHSDGSLATALLFLLLALLLWFWPLCCTVIKEVPPPP---P 358
DB 213 -----SNGIAAIIVLVLLLLGLGLMFWFWPLCKVVKVIRPPPPPPAP 257
QY 359 AESEEE 364
DB 258 KEDEEE 263

RESULT 4
ITAD_HUMAN
ID ITAD_HUMAN STANDARD; PRT; 1162 AA.
AC Q13349; Q15575; Q15576;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-D precursor (Leukointegrin alpha D) (CD11d) (ADB2).
GN ITGAD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=66111956; PubMed=8777714;
RA Van der Vieren M., Le Trong H., Wood C.L., Moore P.F., St John T.,
RA Staunton D.E., Gallatin W.M.;
RT "A novel leukointegrin, alpha d beta 2, binds preferentially to ICAM-
RL 3";
RL Immunity 3:683-690(1995).
RN [2]
RN SEQUENCE OF 1-235 FROM N.A.
RX MEDLINE=20187620; PubMed=10722744;
RA Noti J.D., Johnson A.K., Dillon J.D.;
```

"Structural and functional characterization of the leukocyte integrin gene CD11d. Essential role of Sp1 and Sp3.";

[3] J. Biol. Chem. 275:8959-8969(2000).

SEQUENCE OF 467-537; 571-602; 633-666; 788-834 AND 910-1125 FROM N.A. MEDLINE=96257236; PubMed=8666289;

Wong D.A., Davis E.M., LeBeau M., Springer T.A.;

"Cloning and chromosomal localization of a novel gene-encoding a human beta 2-integrin alpha subunit.";

[4] Gene 171:291-294(1996).

INTERACTION WITH VCAM1.

MEDLINE=99059842; PubMed=9841932;

Grayson M.H., Van der Vieren M., Sterbinsky S.A., Michael Gallatin W., Hoffman P.A., Staunton D.E., Bochner B.S.;

"alpha2beta2 Integrin is expressed on human eosinophils and functions as an alternative ligand for vascular cell adhesion molecule 1 (VCAM-1).";

J. Exp. Med. 188:2187-2191(1998).

[5] INTERACTION WITH VCAM1.

MEDLINE=99370002; PubMed=10438935;

Van der Vieren M., Grove D.T., Hoekstra D., Vazeux R., Hoffman P.A., Grayson M.H., Bochner B.S., Gallatin W.M., Staunton D.E.;

"The leukocyte integrin alpha D beta 2 binds VCAM-1: evidence for a binding interface between I domain and VCAM-1.";

J. Immunol. 163:1984-1990(1999).

[-] FUNCTION: INTEGRIN ALPHA-D/BETA-2 IS A RECEPTOR FOR ICAM3 AND VCAM1. MAY PLAY A ROLE IN THE ATHEROSCLEROTIC PROCESS SUCH AS CLEARING LIPOPROTEINS FROM PLAQUES AND IN PHAGOCYTOSIS OF BLOOD-BORNE PATHOGENS, PARTICULATE MATTER, AND SENESENT ERYTHROCYTES FROM THE BLOOD.

[-] SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-D ASSOCIATES WITH BETA-2.

[-] SUBCELLULAR LOCATION: Type I membrane protein.

[-] TISSUE SPECIFICITY: EXPRESSED MODERATELY ON MYELOMONOCYTIC CELL LINES AND SUBSETS OF PERIPHERAL BLOOD LEUKOCYTES AND STRONGLY ON TISSUE-SPECIALIZED CELLS, INCLUDING MACROPHAGES FOAM CELLS WITHIN ATHEROSCLEROTIC PLAQUES.

[-] DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWF DOMAIN. INTEGRINS WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.

[-] SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.

[-] SIMILARITY: CONTAINS 1 VWF DOMAIN.

[-] SIMILARITY: CONTAINS 7 FG-GAP REPEATS.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; U37028; AAB38547.1; -

EMBL; U40274; AAB60634.1; -

EMBL; U40275; AAB60635.1; -

EMBL; U40276; AAB60636.1; -

EMBL; U40277; AAB60637.1; -

EMBL; U40279; AAB60638.1; -

EMBL; U40278; AAB60638.1; JOINED.

EMBL; AF187881; AAF62875.1; -

HSSP; F11215; IABX.

MIM; 602453; -

InterPro; IPR000413; Integrin\_alpha.

InterPro; IPR002035; VWF.A.

Pfam; PF01839; FG-GAP; 5.

Pfam; PF00357; Integrin\_A; 1.

Pfam; PF00092; vwa; 1.

PRINTS; PR01185; INTEGRINA.

PRINTS; PR00453; VWFADOMAIN.

SMART; SM00191; Int\_alpha; 4.

SMART; SM00327; VWA; 1.

PROSITE; PS00242; INTEGRIN\_ALPHA; 1.

DR		PROSITE: PS50234; VWFA; 1.	
KW	Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;		
KW	Signal; Repeat; Calcium;		
KW	Magnesium.		
FT	SIGNAL	1	17
FT	CHAIN	18	1162
FT	DOMAIN	18	1100
FT	TRANSMEM	1101	1124
FT	DOMAIN	1125	1162
FT	REPEAT	32	85
FT	REPEAT	?	?
FT	DOMAIN	150	332
FT	REPEAT	350	400
FT	REPEAT	401	452
FT	REPEAT	454	516
FT	REPEAT	518	576
FT	REPEAT	581	633
FT	CA_BIND	465	473
FT	CA_BIND	530	538
FT	CA_BIND	593	601
FT	SITE	1127	1131
FT	DISULFID	67	74
FT	DISULFID	106	124
FT	DISULFID	655	710
FT	DISULFID	769	775
FT	DISULFID	846	861
FT	DISULFID	994	1018
FT	DISULFID	1023	1028
FT	CARBOHYD	59	59
FT	CARBOHYD	87	87
FT	CARBOHYD	99	99
FT	CARBOHYD	391	391
FT	CARBOHYD	691	691
FT	CARBOHYD	733	733
FT	CARBOHYD	873	873
FT	CARBOHYD	957	957
FT	CARBOHYD	1046	1046
FT	CONFLICT	500	500
FT	CONFLICT	515	518
FT	CONFLICT	825	825
FT	CONFLICT	984	984
FT	SEQUENCE	1162 AA;	126885 MW; F296A1A35455DD77D CRC64;

[illegible]

DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Integrin alpha-M precursor (Cell surface glycoprotein MAC-1 alpha  
DE subunit) (CR-3 alpha chain) (CD11b) (Leukocyte adhesion receptor Mol1  
DE (Neutrophil adherence receptor).  
DE ITGAM OR CR3A OR CD11b.  
GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=88315033; PubMed=2457584;  
RA Corbi A.L., Kishimoto T.K., Miller L.J., Springer T.A.;  
RT "The human leukocyte adhesion glycoprotein Mac-1 (complement receptor  
RT type 3, CD11b) alpha subunit. Cloning, primary structure, and  
RT relation to the integrins, von Willebrand factor and factor B.";  
RN J. Biol. Chem. 263:12403-12411(1988).  
RN [2]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=88190151; PubMed=2833753;  
RA Arnaut M.A., Remold-O'Donnell E., Pierce M.W., Harris P., Tenen D.G.;  
RT "Molecular cloning of the alpha subunit of human and guinea pig  
RT leukocyte adhesion glycoprotein Mol1: chromosomal localization and  
RT homology to the alpha subunits of integrins.";  
RN Proc. Natl. Acad. Sci. U.S.A. 85:2776-2780(1988).  
RN [3]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=8825215; PubMed=2454931;  
RA Arnaut M.A., Gupta S.K., Pierce M.W., Tenen D.G.;  
RT "Amino acid sequence of the alpha subunit of human leukocyte adhesion  
RT receptor Mol1 (complement receptor type 3).";  
RN J. Cell Biol. 106:2153-2158(1988).  
RN [4]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=93123748; PubMed=8419480;  
RA Fleming J.C., Pahl H.L., Gonzalez D.A., Smith T.F., Tenen D.G.;  
RT "Structural analysis of the CD11b gene and phylogenetic analysis of  
RT the alpha-integrin gene family demonstrate remarkable conservation of  
RT genomic organization and suggest early diversification during  
RT evolution.";  
RN J. Immunol. 150:480-490(1993).  
RN [5]  
RN SEQUENCE OF 9-1153 FROM N.A.  
RX MEDLINE=89098893; PubMed=2563162;  
RA Hickstein D.D., Hickey M.J., Ozols J., Baker D.M., Back A.L.,  
RA Roth G.J.;  
RT "cDNA sequence for the alpha M subunit of the human neutrophil  
RT adherence receptor indicates homology to integrin alpha subunits.";  
RN Proc. Natl. Acad. Sci. U.S.A. 86:257-261(1989).  
RN [6]  
RN SEQUENCE OF 1-9 FROM N.A.  
RX MEDLINE=92073318; PubMed=1683702;  
RA Shelley C.S., Arnaut M.A.;  
RT "The promoter of the CD11b gene directs myeloid-specific and  
RT developmentally regulated expression.";  
RN Proc. Natl. Acad. Sci. U.S.A. 88:10525-10529(1991).  
RN [7]  
RN SEQUENCE OF 1-9 FROM N.A.  
RX MEDLINE=92144986; PubMed=1346576;  
RA Pahl H.L., Rosmarin A.G., Tenen D.G.;  
RT "Characterization of the myeloid-specific CD11b promoter.";  
RN Blood 79:865-870(1992).  
RN [8]  
RN SEQUENCE OF 17-31.  
RX MEDLINE=87076671; PubMed=3539202;  
RA Pierce M.W., Remold-O'Donnell E., Todd R.F. III, Arnaut M.A.;  
RT "N-terminal sequence of human leukocyte glycoprotein Mol1:  
RT conservation across species and homology to platelet IIB/IIIA.";  
RN Biochim. Biophys. Acta 874:368-371(1986).  
RN [9]  
RX X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 148-331.  
RX MEDLINE=95171458; PubMed=7867070;  
RA Lee J.O., Rieu P., Arnaut M.A., Liddington R.;  
RT "Crystal structure of the A domain from the alpha subunit of Integrin  
RT CR3 (CD11b/CD18).";  
RN Cell 80:631-638(1995).  
RN [10]  
RX X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 148-334.  
RX MEDLINE=96363671; PubMed=8747460;  
RA Lee J.O., Banks L.A., Arnaut M.A., Liddington R.C.;  
RT "Two conformations of the integrin A-domain (I-domain): a pathway for  
RT activation?";  
RN Structure 3:1333-1340(1995).  
RN [11]  
RX X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 148-337.  
RX MEDLINE=96363595; PubMed=9687375;  
RA Baldwin E.T., Sarver R.W., Bryant G.L. Jr., Curry K.A.,  
RA Fairbanks M.B., Finzel B.C., Garlick R.L., Heinrichson R.L.,  
RA Horton N.C., Kelley L.L., Mildner A.M., Moon J.B., Mott J.E.,  
RA Mutchler V.T., Tomich C.S., Watenpaugh K.D., Wiley V.H.;  
RT "Cation binding to the integrin CD11b I domain and activation model  
RT assessment.";  
RN Structure 6:923-935(1998).  
RN [12]  
RN 3D-STRUCTURE MODELING OF 17-616.  
RX MEDLINE=98226734; PubMed=9560195;  
RA Oxvig C., Springer T.A.;  
RT "Experimental support for a beta-propeller domain in integrin alpha-  
RT subunits and a calcium binding site on its lower surface.";  
RN Proc. Natl. Acad. Sci. U.S.A. 95:4870-4875(1998).  
RN CC -!- FUNCTION: INTEGRIN ALPHA-M/BETA-2 IS IMPLICATED IN VARIOUS  
RN CC ADHESIVE INTERACTIONS OF MONOCYTES, MACROPHAGES AND GRANULOCYTES  
RN CC AS WELL AS IN MEDIATING THE UPTAKE OF COMPLEMENT-COATED PARTICLES.  
RN CC IT IS IDENTICAL WITH CR-3, THE RECEPTOR FOR THE IC3B FRAGMENT OF  
RN CC PEPTIDE IN C3B. INTEGRIN ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR  
RN CC FIBRINOGEN, FACTOR X AND ICAM1. IT RECOGNIZES P1 AND P2 PEPTIDES  
RN CC OF FIBRINOGEN GAMMA CHAIN.  
RN CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-M  
RN CC ASSOCIATES WITH BETA-2.  
RN CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
RN CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND  
RN CC GRANULOCYTES.  
RN CC -!- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VFMA DOMAIN. INTEGRINS  
RN CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
RN CC -!- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
RN CC -!- SIMILARITY: CONTAINS 1 VFMA DOMAIN.  
RN CC -!- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
RN CC -!- DATABASE: NAME-PROW; NOTE-CD guide CD11b entry;  
RN CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd11b.htm".  
RN CC -----  
RN CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
RN CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
RN CC the European Bioinformatics Institute. There are no restrictions on its  
RN CC use by non-profit institutions as long as its content is in no way  
RN CC modified and this statement is not removed. Usage by and for commercial  
RN CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
RN CC or send an email to license@isb-sib.ch).  
RN CC -----  
RN CC EMBL: J03925; AAA59544.1; -  
RN CC EMBL: M18044; AAA59491.1; -  
RN CC EMBL: J04145; AAA59503.1; -  
RN CC EMBL: S52227; AAB24821.1; -  
RN CC EMBL: S52152; AAB24821.1; JOINED.  
RN CC EMBL: S52153; AAB24821.1; JOINED.  
RN CC EMBL: S52154; AAB24821.1; JOINED.  
RN CC EMBL: S52155; AAB24821.1; JOINED.  
RN CC EMBL: S52157; AAB24821.1; JOINED.  
RN CC EMBL: S52159; AAB24821.1; JOINED.  
RN CC EMBL: S52161; AAB24821.1; JOINED.  
RN CC EMBL: S52164; AAB24821.1; JOINED.  
RN CC EMBL: S52165; AAB24821.1; JOINED.  
RN CC EMBL: S52167; AAB24821.1; JOINED.

DR	EMBL	S52169	AAB24821.1	JOINED.
DR	EMBL	S52170	AAB24821.1	JOINED.
DR	EMBL	S52173	AAB24821.1	JOINED.
DR	EMBL	S52174	AAB24821.1	JOINED.
DR	EMBL	S52180	AAB24821.1	JOINED.
DR	EMBL	S52181	AAB24821.1	JOINED.
DR	EMBL	S52184	AAB24821.1	JOINED.
DR	EMBL	S52189	AAB24821.1	JOINED.
DR	EMBL	S52191	AAB24821.1	JOINED.
DR	EMBL	S52192	AAB24821.1	JOINED.
DR	EMBL	S52203	AAB24821.1	JOINED.
DR	EMBL	S52212	AAB24821.1	JOINED.
DR	EMBL	S52213	AAB24821.1	JOINED.
DR	EMBL	S52216	AAB24821.1	JOINED.
DR	EMBL	S52219	AAB24821.1	JOINED.
DR	EMBL	S52220	AAB24821.1	JOINED.
DR	EMBL	S52221	AAB24821.1	JOINED.
DR	EMBL	S52222	AAB24821.1	JOINED.
DR	EMBL	S52226	AAB24821.1	JOINED.
DR	EMBL	M76724	AAA58410.1	-
DR	EMBL	M84477	AAA51960.1	-
DR	PIR	A31108	RWHULB	-
DR	PIR	A26091	A26091	-
DR	PDB	1A8X	17-JUN-98	-
DR	PDB	1BHO	18-NOV-98	-
DR	PDB	1BHQ	18-NOV-98	-
DR	PDB	1IDN	25-NOV-98	-
DR	PDB	1IDO	01-AUG-96	-
DR	PDB	1JLM	11-JAN-97	-
DR	CarbBank	CCSD:33581	-	-
DR	CarbBank	CCSD:33582	-	-
DR	CarbBank	CCSD:33584	-	-
DR	MIM	120980	-	-
DR	InterPro	IPR000413	Integrin_alpha	-
DR	InterPro	IPR002035	WVFA	-
DR	Pfam	PF01839	FG-GAP; 5	-
DR	Pfam	PF00357	Integrin_A; 1	-
DR	Pfam	PF00092	vwa; 1	-
DR	PRINTS	PRO1185	INTEGRINA	-
DR	PRINTS	PRO0453	WVFADOMAIN	-
DR	SMART	SM00191	Int_alpha; 4	-
DR	SMART	SM00327	WVA; 1	-
DR	PROSITE	PS00242	INTEGRIN_ALPHA; 1	-
DR	PROSITE	PS0234	WVFA; 1	-
DR	Integrin	Cell adhesion; Receptor; Glycoprotein; Transmembrane;	-	-
KW	Signal	3D-structure; Repeat; Magnesium; Calcium.	-	-
FT	SIGNAL	1 16	-	-
FT	CHAIN	17 1152	INTEGRIN ALPHA-M.	-
FT	DOMAIN	17 1104	EXTRACELLULAR (POTENTIAL).	-
FT	TRANSMEM	1105 1128	POTENTIAL.	-
FT	DOMAIN	1129 1152	CYTOPLASMIC (POTENTIAL).	-
FT	REPEAT	31 84	FG-GAP 1.	-
Query Match 7.6%; Score 145.5; DB 1; Length 1152;				
Best Local Similarity 26.2%; Pred. No. 0.0006;				
Matches 60; Conservative 45; Mismatches 81; Indels 43; Gaps 12;				
Qy	44	DLVFILDKSGVL-HHWNEIYFEQLAHKFTISPOLRMSFIVFSTRGTTMLKTED----	98	
Db	150	DIAFLIDGSGSIIPDFRRMKFVST-----VMEQLKSKTLFS-----LMOYSEEFRIH 199		
Qy	99	-----RQIQGLLEOKVLPGDVTYMHGFERASQIYYENRQGYRTA-SVITALT 149		
Db	200	FTFKEFQNNPNRSLVKPIQLL--GRHTATGIRKVVRLFNITNGARKNAFKILVIT 257		
Qy	150	DELGHDLFFVSE--REANSRDLGAIVCGVKDFNETQLAR-----TADS--KDHVFP 200		
Db	258	DGEFGDPLGYEDVIPADRE---GVRIYVIGVDAPFRSEKRSQELNTIASKPPRDHFVQ 314		
Qy	201	VNDGFOALQGIHLSILKKSCIEILAAPSTICAGESFOVVVRNGRHA 249		
Db	315	VNN-FAELKTIONLREK----IFAIEGTQTGSSSSFEHMSQEGFSAA 358		

RESULT	6
ID	ITAX_HUMAN
STANDARD	PRT; 1163 AA.
AC	P20702;
DT	01-FEB-1991 (Rel. 17, Created)
DT	01-FEB-1991 (Rel. 17, Last sequence update)
DT	01-MAR-2002 (Rel. 41, Last annotation update)
DE	Integrin alpha-X precursor (Leukocyte adhesion glycoprotein p150,95 alpha chain) (Leukocyte adhesion receptor p150,95) (CD11c) (Leu M5).
GN	ITGAX OR CD11c.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A. PubMed=3327687;
RX	MEDLINE=86166645; PubMed=2303426;
RA	Corbi A.L., Miller L.J., O'Connor K., Larson R.S., Springer T.A.;
RT	"CDNA cloning and complete primary structure of the alpha subunit of a leukocyte adhesion glycoprotein, p150,95.";
RT	EMBO J. 6:4023-4028(1987).
RL	[2]
RN	SEQUENCE FROM N.A.
RP	MEDLINE=90153906; PubMed=2303426;
RX	Corbi A.L., Garcia-Aguilar J., Springer T.A.;
RT	"Genomic structure of an integrin alpha subunit, the leukocyte p150,95 molecule.";
RT	J. Biol. Chem. 265:2782-2788(1990).
RL	[3]
RN	ERRATUM.
RA	Corbi A.L., Garcia-Aguilar J., Springer T.A.;
RL	J. Biol. Chem. 265:12750-12751(1990).
RL	[4]
RN	SEQUENCE OF 20-43.
RP	MEDLINE=87167596; PubMed=3549901;
RX	Miller L.J., Wiebe M., Springer T.A.;
RT	"Purification and alpha subunit N-terminal sequences of human Mac-1 and p150,95 leukocyte adhesion proteins.";
RT	J. Immunol. 138:2381-2383(1987).
RL	[5]
RN	FUNCTION: INTEGRIN ALPHA-X/BETA-2 IS A RECEPTOR FOR FIBRINOGEN. IT RECOGNIZES THE SEQUENCE G-P-R IN FIBRINOGEN. IT MEDIATES CELL-CELL INTERACTION DURING INFLAMMATORY RESPONSES. IT IS ESPECIALLY IMPORTANT IN MONOCYTE ADHESION AND CHEMOTAXIS.
CC	-1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-X ASSOCIATES WITH BETA-2.
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.
CC	-1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND GRANULOCYTES.
CC	-1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A WVFA DOMAIN. INTEGRINS WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC	-1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC	-1- SIMILARITY: CONTAINS 1 WVFA DOMAIN.
CC	-1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC	-1- DATABASE: NAME=PROW; NOTE=CD guide CD11c entry; WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd11c.htm".
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; M81695; AAA59180.1; -
DR	EMBL; Y00093; CAA68283.1; -
DR	EMBL; M29165; -; NOT ANNOTATED CDS.
DR	EMBL; M29487; AAA51620.1; ALT_SEQ.
DR	EMBL; M29482; AAA51620.1; JOINED.
DR	EMBL; M29483; AAA51620.1; JOINED.
DR	EMBL; M29484; AAA51620.1; JOINED.
DR	EMBL; M29485; AAA51620.1; JOINED.





-1- TISSUE SPECIFICITY: TYPE XII COLLAGEN IS PRESENT IN TENDONS, LIGAMENTS, PERICHONDRUM, AND PERIOSTEUM, ALL DENSE CONNECTIVE TISSUES CONTAINING TYPE I COLLAGEN.

-1- DOMAIN: THIS SEQUENCE DEFINES FIVE DISTINCT DOMAINS, TWO TRIPLE-HELICAL DOMAINS (COL1 AND COL2) AND THREE NONTRIPLE-HELICAL DOMAINS (NC1, NC2, AND NC3).

-1- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT EACH END.

-1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

-1- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY).

-1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH INTERRUPTED HELICES (FACIT) FAMILY.

-1- SIMILARITY: CONTAINS 4 VWFA DOMAINS.

-1- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: D00824; BAA00701.1; -  
EMBL: X61024; CAA43358.1; -  
EMBL: M17375; AAA48718.1; -  
EMBL: J05137; AAA48635.1; -  
EMBL: X67327; CAA47744.1; -  
PIR: A28037; A28037.  
PIR: A34485; A34485.  
HSSP: P17301; 1A0X.  
InterPro: IPR000087; Collagen.  
InterPro: IPR003961; FN III.  
InterPro: IPR003962; FnIII\_repeat.  
InterPro: IPR003129; TSPN.  
InterPro: IPR002035; VWFA.  
Pfam: PF01391; Collagen: 4.  
Pfam: PF00041; fn3: 17.  
Pfam: PF02210; TSPN: 1.  
Pfam: PF00092; vwa: 4.  
PRINTS: PR00014; FNYPEIII.  
PRINTS: PR00453; VWFADOMAIN.  
SMART: SM00060; FN3: 17.  
SMART: SM00210; TSPN: 1.  
SMART: SM00327; VWA: 4.  
PROSITE: PS02034; VWFA: 4.  
Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.  
KW SIGNAL  
KW 1 24  
FT CHAIN 25 3124 COLLAGEN ALPHA 1(XII) CHAIN.  
FT DOMAIN 25 114 FIBRONECTIN TYPE-III 1.  
FT DOMAIN 139 311 VWFA 1.  
FT DOMAIN 332 425 FIBRONECTIN TYPE-III 2.  
FT DOMAIN 439 615 VWFA 2.  
FT DOMAIN 629 720 FIBRONECTIN TYPE-III 3.  
FT DOMAIN 721 811 FIBRONECTIN TYPE-III 4.  
FT DOMAIN 812 904 FIBRONECTIN TYPE-III 5.  
FT DOMAIN 905 998 FIBRONECTIN TYPE-III 6.  
FT DOMAIN 999 1085 FIBRONECTIN TYPE-III 7.  
FT DOMAIN 1086 1178 FIBRONECTIN TYPE-III 8.  
FT DOMAIN 1199 1371 VWFA 3.  
FT DOMAIN 1384 1473 FIBRONECTIN TYPE-III 9.  
FT DOMAIN 1474 1565 FIBRONECTIN TYPE-III 10.  
FT DOMAIN 1566 1654 FIBRONECTIN TYPE-III 11.  
FT DOMAIN 1655 1755 FIBRONECTIN TYPE-III 12.  
FT DOMAIN 1756 1846 FIBRONECTIN TYPE-III 13.  
FT DOMAIN 1847 1936 FIBRONECTIN TYPE-III 14.  
FT DOMAIN 1937 2027 FIBRONECTIN TYPE-III 15.  
FT DOMAIN 2028 2118 FIBRONECTIN TYPE-III 16.  
FT DOMAIN 2119 2206 FIBRONECTIN TYPE-III 17.  
FT DOMAIN 2207 2295 FIBRONECTIN TYPE-III 18.  
FT DOMAIN 2327 2500 VWFA 4.

FT	DOMAIN	2455	2750	NONHELICAL REGION (NC3)
FT	DOMAIN	2751	2902	TRIPLE-HELICAL REGION (COL2)
FT				WITH 1 IMPERFECTION.
FT	DOMAIN	2903	2945	NONHELICAL REGION (NC2).
FT	DOMAIN	2946	3048	TRIPLE-HELICAL REGION (COL1)
FT				WITH 2 IMPERFECTIONS.
FT	DOMAIN	3049	3124	NONHELICAL REGION (NC1).
FT	DOMAIN	3086	3096	ASP/GLU-RICH (ACIDIC).
FT	DOMAIN	3111	3123	ARG/LYS-RICH (BASIC).
FT	BINDING	797	797	TO CHONDROITIN SULFATE (POTENTIAL).
FT	BINDING	890	890	TO CHONDROITIN SULFATE (POTENTIAL).
FT	BINDING	981	981	TO CHONDROITIN SULFATE (POTENTIAL).
FT	SITE	2438	2440	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	2899	2901	CELL ATTACHMENT SITE (POTENTIAL).
FT	CARBOHYD	32	32	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1006	1006	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1032	1032	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1044	1044	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1512	1512	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1767	1767	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2210	2210	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2273	2273	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2532	2532	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2683	2683	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARSPLIC	25	1188	MISSING (IN SHORT ISOFORM).
FT	CONFLICT	1258	1258	T -> S (IN REF. 4).
FT	CONFLICT	1264	1264	D -> E (IN REF. 4).
FT	CONFLICT	2759	2759	P -> A (IN REF. 2).
FT	CONFLICT	2803	2803	L -> F (IN REF. 2).
FT	CONFLICT	2977	2977	V -> F (IN REF. 2).
FT	CONFLICT	3075	3076	QP -> AG (IN REF. 3).
FT	CONFLICT	3077	3076	
FT	SEQUENCE	3124	340578	MM; 094285AFE7F346CF CRC64;
FT	SEQUENCE	3124	AA;	

Query Match 7.5%; Score 143; DB 1; Length 3124;

Best Local Similarity 26.1%; Pred. No. 0.0035;

Matches 65; Conservative 51; Mismatches 97; Indels 36; Gaps

QY	44	DLYLFDKGSQV-LHHWNEIYFVEQLAHKF-ISP-QLRMSFIVFSTGTTLMKLTEDRE	100
DB	439	DVFLVDGYSIGIANFVKVRAFLVLFVKSFEISPRKVOISLVQYSR--DPHMEFSLNRY	496
QY	101	QIROGLELQKVLIP--CGDYYMHGEPERASEQIYYENROGYR--TASVLIATLDGDELHD	156
DB	497	NRVKDIIQAINTFPYRGSGTNTGKAMTYVREKVVFTSK-GSRPNVPRVMLITDCK-SSD	554
QY	157	LFFFYSERANRSRLDGAIVCYGVGFENETQLARIAD--SKDHVPVNDGFOALQIHS	214
DB	555	AF---KEPAIKLRDADVEIFAVGVKDAVRTELEAIASPPAETHYVYVED-FDAFORISPE	610
QY	215	ILKKSCEILAAEPSTTCAGESFOVYVRNGFRHARW---DRVLGCFKINDSVLTNEKP	271
DB	611	LQTSVCLRI-----EQELAAIRKKSYPVPAKNNWFSDTSDSFVKSWSAAGSE--	657
QY	272	FSVEDTYLL	280
DB	658	---EKSYLI	663

RESULT 8

ITAM\_MOUSE

ID ITAM\_MOUSE STANDARD; PRT; 1153 AA.

AC	P05555;	
DT	01-NOV-1988 (Rel. 09, Created)	
DT	01-FEB-1991 (Rel. 17, Last sequence update)	
DT	01-WAR-2002 (Rel. 41, Last annotation update)	
DE	Integrin alpha-M precursor (Cell surface glycoprotein MAC-1 alpha subunit) (CR-3 alpha chain) (CD11b) (Leukocyte adhesion receptor MO1)	
GN	ITGAM.	
OC	Mus musculus (Mouse).	
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_Taxid=10090;	

[1] SEQUENCE FROM N.A.  
 RP MEDLINE-8312584; PubMed=3044779;  
 RA Pyela R.;  
 RT "Amino acid sequence of the murine Mac-1 alpha chain reveals homology  
 with the integrin family and an additional domain related to von  
 Willebrand factor.";  
 RL EMBO J. 7:1371-1378(1988).  
 [2]  
 RP SEQUENCE OF 11-45 FROM N.A.  
 RC STRAIN-BALB/C; TISSUE=Spleen;  
 RA MEDLINE-86287312; PubMed=2942940;  
 RA Sastre L., Roman J.M., Teplow D.B., Dreyer W.J., Gee C.E.,  
 RA Larson R.S., Roberts T.M., Springer T.A.;  
 RT "A partial genomic DNA clone for the alpha subunit of the mouse  
 complement receptor type 3 and cellular adhesion molecule Mac-1.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5644-5648(1986).  
 [3]  
 RP SEQUENCE OF 17-28.  
 RX MEDLINE-8518276; PubMed=3887182;  
 RA Springer T.A., Teplow D.B., Dreyer W.J.;  
 RT "Sequence homology of the LFA-1 and Mac-1 leukocyte adhesion  
 glycoproteins and unexpected relation to leukocyte interferon.";  
 RL Nature 314:540-542(1985).  
 CC -!- FUNCTION: INTEGRIN ALPHA-M/BETA-2 IS IMPLICATED IN VARIOUS  
 ADHESIVE INTERACTIONS OF MONOCYTES, MACROPHAGES AND GRANULOCYTES  
 AS WELL AS IN MEDIATING THE UPTAKE OF COMPLEMENT-COATED PARTICLES.  
 IT IS IDENTICAL WITH CR-3, THE RECEPTOR FOR THE IC3B FRAGMENT OF  
 THE THIRD COMPLEMENT COMPONENT. IT PROBABLY RECOGNIZES THE R-G-D  
 PEPTIDE IN C3B. INTEGRIN ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR  
 FIBRINOGEN, FACTOR X AND ICAM1. IT RECOGNIZES P1 AND P2 PEPTIDES  
 OF FIBRINOGEN GAMMA CHAIN. ALPHA-M/BETA-2 PLAY A CRITICAL ROLE IN  
 MAST CELL DEVELOPMENT AND IN IMMUNE COMPLEX-MEDIATED  
 GLOMERULONEPHRITIS. MICE EXPRESSING A NULL MUTATION OF THE ALPHA-M  
 SUBUNIT GENE DEMONSTRATE INCREASE IN NEUTROPHIL ACCUMULATION, IN  
 RESPONSE TO AN IMPAIRED DEGRANULATION AND PHAGOCYTOSIS, EVENTS THAT  
 APPARENTLY ACCELERATE APOPTOSIS IN NEUTROPHILS. THESE MICE DEVELOP  
 OBESITY.  
 CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN. ALPHA-M  
 ASSOCIATES WITH BETA-2.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND  
 GRANULOCYTES.  
 CC -!- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS  
 WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
 CC -!- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
 CC -!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X07640; CAA30479.1; -  
 DR EMBL: M14293; AAA39484.1; -  
 DR PIR: S00551; S00551.  
 DR HSP: P11215; IABX.  
 DR MGI: 96607; Itgam.  
 DR InterPro: IPR000413; Integrin\_alpha.  
 DR InterPro: IPR002035; VWFA.  
 DR Pfam: PF01839; FG-GAP; 5.  
 DR Pfam: PF00357; integrin\_A; 1.  
 DR Pfam: PF00092; vwa; 1.  
 DR PRINTS: PR01185; INTEGRINA.  
 DR PRINTS: PR00453; VWFADOMAIN.  
 DR SMART: SM00191; Int\_alpha; 4.  
 DR SMART: SM00327; VWA; 1.  
 DR PROSITE: PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE: PS0234; VWFA; 1.

KW Integrin: Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
 KW Signal: Calcium; Repeat.  
 FT SIGNAL 1 16  
 FT CHAIN 17 1153 INTEGRIN ALPHA-M.  
 FT DOMAIN 17 1105 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1106 1129 POTENTIAL.  
 FT DOMAIN 1130 1153 CYTOPLASMIC (POTENTIAL).  
 FT REPEAT 31 84 FG-GAP 1.  
 FT REPEAT ? ? FG-GAP 2.  
 FT DOMAIN 164 350 VWFA.  
 FT REPEAT 337 400 FG-GAP 3.  
 FT REPEAT 401 452 FG-GAP 4.  
 FT REPEAT 454 515 FG-GAP 5.  
 FT REPEAT 517 575 FG-GAP 6.  
 FT REPEAT 580 632 FG-GAP 7.  
 FT CALBIND 465 473 POTENTIAL.  
 FT CA\_BIND 529 537 POTENTIAL.  
 FT CA\_BIND 592 600 POTENTIAL.  
 FT SITE 1132 1136 GFFKR MOTIF.  
 FT DISULFID 66 73 BY SIMILARITY.  
 FT DISULFID 105 123 BY SIMILARITY.  
 FT DISULFID 654 711 BY SIMILARITY.  
 FT DISULFID 770 776 BY SIMILARITY.  
 FT DISULFID 999 1023 BY SIMILARITY.  
 FT DISULFID 1028 1033 BY SIMILARITY.  
 FT CARBOHYD 58 58 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 391 391 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 596 596 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 734 734 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 772 772 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 801 801 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 881 881 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 907 907 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 941 941 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 980 980 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 994 994 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1045 1045 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1051 1051 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1076 1076 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 1153 AA; 127480 MW; 178DB988AECB0343 CRC64;  
 Query Match 7.4%; Score 141; DB 1; Length 1153;  
 Best Local Similarity 24.2%; Pred. No. 0.0014;  
 Matches 71; Conservative 50; Mismatches 96; Indels 76; Gaps 16;  
 QY 44 DLYFILDKSGSVLHHWNEIYFVEQLAHKFIS---PQLRMSEIVFSTGTTLMLKLTED-- 98  
 DB 150 DIVFLIDGSGSI-----NNIDF---QMKKEFVSTVMEQFKSKTLFS-----LMQYSDEFR 197  
 QY 99 -----REQIRQGLEELQKVL-----GGDTVMHEGFERASEQIYVE-NRQGYRTASVIALT 149  
 DB 198 IHFTFNDFKRNPSRSHVSPKQLNGRTKTASGIRKRVRELFHKTKNGARENAKILVVT 257  
 QY 150 DGEHLEDLFYSE--REANRSDDLGAIVYCVGVKD--FNETQLARIAD-----SKDHVPP 200  
 DB 258 DGEKFGDPLDYKDVPEADRA---GVIRVIGVGNFNRKPSRREDLTTASKPAGEHVQ 314  
 QY 201 VNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQVWVRNGRPHARNVDRLVCSFK 260  
 DB 315 V-DNFEALNTIQNLQEK-----IFALEGQTGTSTSSFEHEMSQEGF----- 355  
 QY 261 INDSVTLNKPKPSVEDTYLLCPAPILKEVG----MKAALQVSMNDGLSFISS 309  
 DB 356 -SASITSN-----GPLLGSVGSFDMWAGAFLYTSKDKVTFINTT 393  
 RESULT 9  
 CALC\_MOUSE  
 ID CALC\_MOUSE STANDARD; PRT: 3067 AA.  
 AC Q60847;

DR PROSITE: PS50234; VWFA; 4.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.  
FT SIGNAL 1 24  
FT CHAIN 25 3067  
FT DOMAIN 25 114  
FT DOMAIN 140 316  
FT DOMAIN 332 423  
FT DOMAIN 444 620  
FT DOMAIN 634 725  
FT DOMAIN 726 816  
FT DOMAIN 817 907  
FT DOMAIN 908 1002  
FT DOMAIN 1003 1089  
FT DOMAIN 1090 1182  
FT DOMAIN 1203 1375  
FT DOMAIN 1388 1474  
FT DOMAIN 1480 1568  
FT DOMAIN 1569 1652  
FT DOMAIN 1659 1757  
FT DOMAIN 1758 1848  
FT DOMAIN 1849 1938  
FT DOMAIN 1939 2029  
FT DOMAIN 2030 2120  
FT DOMAIN 2121 2208  
FT DOMAIN 2209 2297  
FT DOMAIN 2329 2501  
FT DOMAIN 2456 2751  
FT DOMAIN 2752 2899  
FT DOMAIN 2900 2942  
FT DOMAIN 2943 3045  
FT DOMAIN 3046 3064  
FT BINDING 802 802  
FT BINDING 893 893  
FT BINDING 985 985  
FT SITE 866 868  
FT SITE 2784 2786  
FT SITE 2896 2898  
FT MOD\_RES 2945 2945  
FT MOD\_RES 2948 2948  
FT MOD\_RES 2951 2951  
FT MOD\_RES 2960 2960  
FT MOD\_RES 2966 2966  
FT MOD\_RES 2969 2969  
FT MOD\_RES 2972 2972  
FT MOD\_RES 2984 2984  
FT MOD\_RES 3001 3001  
FT MOD\_RES 3004 3004  
FT MOD\_RES 3015 3015  
FT MOD\_RES 3024 3024  
FT MOD\_RES 3027 3027  
FT MOD\_RES 3030 3030  
FT MOD\_RES 869 872  
FT CARBOHYD 704 704  
FT CARBOHYD 1769 1769  
FT CARBOHYD 2212 2212  
FT CARBOHYD 2533 2533  
FT CARBOHYD 2684 2684  
FT VARSPLIC 25 1190  
SQ SEQUENCE 3067 AA; 334094 MW; 3FB5DEFB8A2CDB95 CRC64;  
  
Query Match 7.3%; Score 139; DB 1; Length 3067;  
Best Local Similarity 30.3%; Pred. No. 0.0072;  
Matches 60; Conservative 37; Mismatches 75; Indels 26; Gaps 12;  
  
QY 44 DLYFILDKSGSV-LHHWNIYVFEOLAHKF-ISP-QLRMSFIVFSTRTTLMKLTEDRE 100  
DB 444 DIVFLVDGSIISGTANFVKVRAFLVLAKEFSPNVOISLVQVSRDPHTFETLKEFN 503  
QY 101 QIRQGLEELQKVL-----EGDGYMHEGFERASEQIYENROGYRT--ASVIALTDGEL 153

```

Db 504 -----VEDIIKAIINTEPYRGSTNGRAMTVREKIEFVKNK-GSRSNVPKVMILITDGE- 556
QY 154 HEDLFFYSERANRDLGALVYGVKDNETOLARIAD--SKDHVPVNDGFGQALQGI 211
Db 557 SSDAF---RPAIKLRSDVEIFAVGVKDAVRSELEAIASPAETHFTVED-FDAFQRI 612
QY 212 IHSILKKSCTEI---LAA 226
Db 613 SFVLTSICLRIGELAA 630

RESULT 10
CO2_MOUSE
ID CO2_MOUSE STANDARD; PRT; 760 AA.
AC P21180;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Complement C2 precursor (EC 3.4.21.43) (C3/C5 convertase).
GN C2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91035430; PubMed=2229060;
RT Ishikawa N., Nonaka M., Wetzel R.A., Colten H.R.;
RT "Murine complement C2 and factor B, a serine protease, and cDNA cloning reveals
RT different mechanisms for multiple transcripts of C2 and B.";
RL J. Biol. Chem. 265:19040-19046(1990).
CC -1- FUNCTION: COMPONENT C2 WHICH IS PART OF THE CLASSICAL PATHWAY OF
CC THE COMPLEMENT SYSTEM IS CLEAVED BY ACTIVATED FACTOR C1 INTO TWO
CC FRAGMENTS: C2B AND C2A. C2A, A SERINE PROTEASE, THEN COMBINES WITH
CC COMPLEMENT FACTOR 4B TO GENERATE THE C3 OR C5 CONVERTASE.
CC -1- CATALYTIC ACTIVITY: CLEAVES C3 IN THE ALPHA-CHAIN TO YIELD C3A AND
CC C3B. CLEAVES C5 IN THE ALPHA-CHAIN TO YIELD C5A AND C5B. BOTH
CC CLEAVAGES TAKE PLACE AT THE C-TERMINAL OF AN ARGININE RESIDUE.
CC -1- MISCELLANEOUS: C2 IS A MAJOR HISTOCOMPATIBILITY COMPLEX CLASS-III
CC PROTEIN.
CC -1- SIMILARITY: WITH COMPLEMENT FACTOR B.
CC -1- SIMILARITY: CONTAINS 3 SUSHI (SCR) DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M60579; AAA37380.1; .
DR EMBL; M60563; AAA37380.1; JOINED.
DR EMBL; M60564; AAA37380.1; JOINED.
DR EMBL; M60565; AAA37380.1; JOINED.
DR EMBL; M60566; AAA37380.1; JOINED.
DR EMBL; M60567; AAA37380.1; JOINED.
DR EMBL; M60568; AAA37380.1; JOINED.
DR EMBL; M60569; AAA37380.1; JOINED.
DR EMBL; M60570; AAA37380.1; JOINED.
DR EMBL; M60571; AAA37380.1; JOINED.
DR EMBL; M60572; AAA37380.1; JOINED.
DR EMBL; M60573; AAA37380.1; JOINED.
DR EMBL; M60574; AAA37380.1; JOINED.
DR EMBL; M60575; AAA37380.1; JOINED.
DR EMBL; M60576; AAA37380.1; JOINED.
DR EMBL; M60577; AAA37380.1; JOINED.
DR EMBL; M60578; AAA37380.1; JOINED.
DR HSP; P00757; LSGF.
DR MEROPS; S01.194; .
DR MGD; MGI:88226; C2.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR001254; Trypsin.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00084; sushi; 2.
DR Pfam; PF00089; trypsin; 2.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00032; CCP; 2.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PROSITE; PS00234; VWFA; 1.
KW Complement pathway; Plasma; Glycoprotein; Hydrolase; Serine protease;
KW Signal; Repeat; Sushi; Alternative splicing.
FT SIGNAL 1 18
FT CHAIN 19 760 COMPLEMENT C2.
FT CHAIN 19 250 COMPLEMENT C2B FRAGMENT.
FT CHAIN 251 760 COMPLEMENT C2A FRAGMENT.
FT DOMAIN 21 90 SUSHI 1.
FT DOMAIN 93 150 SUSHI 2.
FT DOMAIN 155 211 SUSHI 3.
FT DOMAIN 261 459 VWFA.
FT DOMAIN 473 760 SERINE PROTEASE.
FT ACT_SITE 514 514 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 570 570 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 689 689 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 22 62 BY SIMILARITY.
FT DISULFID 49 89 BY SIMILARITY.
FT DISULFID 94 136 BY SIMILARITY.
FT DISULFID 122 149 BY SIMILARITY.
FT DISULFID 156 197 BY SIMILARITY.
FT DISULFID 182 210 BY SIMILARITY.
FT CARBOHYD 27 27 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 32 32 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 117 117 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 340 340 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 474 474 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 663 663 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPIC 606 612 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 760 AA; 84726 MW; D7DE9DEF4DBA48D9 CRC64;

```

Query Match 6.88; Score 131; DB 1; Length 760;  
 Best Local Similarity 23.28; Pred. No. 0.0054;  
 Matches 74; Conservative 56; Mismatches 115; Indels 74; Gaps 16;

Fri Aug 9 10:57:00 2002

```

QY 41 GGDFLYFILDKSGSVLHHWNEY-YVEOLAHKFISPOLR--MSFIVSTRGTTLMKLTE 97
DB 258 GHLNLYLLDASQSVTEKDFDKKSAELMWVERIFSEVNVTVAILTFASQPKTIMSL 317
QY 98 DREIQRQLEELQKVLPGDQTYHGEFASQIYVE-----NRQGYRTAS- 143
DB 318 ERSQ-----DVEVTISLDSASYKDHNATGANTYEVILRVYSMOTQMDRLGHWTSW 371
QY 144 -----VIIATDGLHEDLFFYSREANRSRLGAT-----YVCVGV-----KD 182
DB 372 KEIRTHIILLTDGK--SNMGDSPKKAATRIRELLSTIEQNRDDYLDIYAIGVGKLDVDWKE 429
QY 183 FNETQIARADSKDHVPVNDQFALQGIHSLKSKSCIEILAAEPSTTCAGESFOVVYR 242
DB 430 LNE--LGSKKDGERHAFILQDA-KALQOIFEHMLDSKL-----TDTICG-----V 472
QY 243 GNGFRHARNVDV--LCSFKINDSVTLNEKPFVSDVETLLCPAPILKEVGMK--AALQVYS 298
DB 473 GNSANASQDERTPWQVTFKPSKETCQGS--LISDQWVLTAAHCFDIQMEDHILWRVN 530
QY 299 MND-----GLSFISSSVII 312
DB 531 VGDPTSQHGKEFLVEDVII 549

RESULT 11
CA1C_HUMAN STANDARD; PRT: 3063 AA.
AC Q99715; Q99716;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(XII) chain precursor.
GN COL12A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (LONG AND SHORT FORM).
RP MEDLINE=97288521; PubMed=9143499;
RA Gerecke D.R., Olson P.F., Koch M., Knoll J.H.M., Taylor R.,
RA Hudson D.L., Champilaud M.-F., Olsen B.R., Burgeson R.E.;
RT "Complete primary structure of two splice variants of collagen xii,
RT and assignment of alpha 1(XII) collagen (COL12A1), alpha 1(IX)
RT collagen (COL9A1), and alpha 1(XIX) collagen (COL19A1) to human
RT chromosome 6q12-q13."
RL Chromosomes 41:236-242(1997).
CC -!- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-
CC CONTAINING FIBRILS, THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE
CC SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE
CC LOCALIZED IN THE PERIFIBRILLAR MATRIX (BY SIMILARITY).
CC -!- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 KDA OF
CC NONTRIPLE-HELICAL SEQUENCES.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE FINAL TISSUE
CC FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OF EITHER THE LONGER
CC OR THE SHORTER ISOFORM OR ANY COMBINATION OF LONG AND SHORT
CC ISOFORM CHAINS. ONLY THE LONG VARIANT IS A PROTEOLYCAN.
CC -!- TISSUE SPECIFICITY: FOUND IN COLLAGEN I-CONTAINING TISSUES: BOTH
CC SHORT AND LONG ISOFORMS APPEAR IN ANNIION, CHORION, SKELETAL
CC MUSCLE, SMALL INTESTINE, AND IN CELL CULTURE OF DERMAL
CC FIBROBLASTS, KERATINOCYTES, AND ENDOTHELIAL CELLS. ONLY THE SHORT
CC ISOFORM IS FOUND IN LUNG, PLACENTA, KIDNEY, AND A SQUAMOUS CELL
CC CARCINOMA CELL LINE.
CC -!- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT
CC EACH END (BY SIMILARITY).
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY
CC SIMILARITY).
CC -!- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH

```

```

CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -!- SIMILARITY: CONTAINS 4 VWFA DOMAINS.
CC -!- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U73778; AAC51244.1; -.
CC EMBL; U73779; AAD40483.1; -.
CC HSSP; P02751; 1TTF.
CC MIM; 120320; -.
CC InterPro; IPR000087; Collagen.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR003962; FNIII_repeat.
CC InterPro; IPR003129; TSPN.
CC InterPro; IPR002035; VWFA.
CC Pfam; PF01391; Collagen; 4.
CC Pfam; PF00041; fn3; 18.
CC Pfam; PF02210; TSPN; 1.
CC Pfam; PF00092; vwa; 4.
CC PRINTS; PR00014; FNYPEIII.
CC PRINTS; PR00453; VWFADOMAIN.
CC SMART; SM00060; FN3; 17.
CC SMART; SM00210; TSPN; 1.
CC SMART; SM00327; VWA; 4.
CC PROSITE; PS50234; VWFA; 4.
CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.
KW SIGNAL
FT CHAIN 25 3063
FT COLLAGEN ALPHA 1(XII) CHAIN.
FT FIBRONECTIN TYPE-III 1.
FT DOMAIN 25 114
FT VWFA 1.
FT DOMAIN 140 316
FT FIBRONECTIN TYPE-III 2.
FT DOMAIN 333 426
FT VWFA 2.
FT DOMAIN 440 616
FT FIBRONECTIN TYPE-III 3.
FT DOMAIN 630 721
FT FIBRONECTIN TYPE-III 4.
FT DOMAIN 722 812
FT FIBRONECTIN TYPE-III 5.
FT DOMAIN 813 903
FT FIBRONECTIN TYPE-III 6.
FT DOMAIN 904 998
FT FIBRONECTIN TYPE-III 7.
FT DOMAIN 999 1085
FT FIBRONECTIN TYPE-III 8.
FT DOMAIN 1086 1178
FT VWFA 3.
FT DOMAIN 1199 1371
FT FIBRONECTIN TYPE-III 9.
FT DOMAIN 1384 1473
FT FIBRONECTIN TYPE-III 10.
FT DOMAIN 1474 1564
FT FIBRONECTIN TYPE-III 11.
FT DOMAIN 1565 1652
FT FIBRONECTIN TYPE-III 12.
FT DOMAIN 1654 1751
FT FIBRONECTIN TYPE-III 13.
FT DOMAIN 1752 1842
FT FIBRONECTIN TYPE-III 14.
FT DOMAIN 1843 1932
FT FIBRONECTIN TYPE-III 15.
FT DOMAIN 1933 2023
FT FIBRONECTIN TYPE-III 16.
FT DOMAIN 2024 2114
FT FIBRONECTIN TYPE-III 17.
FT DOMAIN 2115 2202
FT FIBRONECTIN TYPE-III 18.
FT DOMAIN 2203 2291
FT VWFA 4.
FT DOMAIN 2323 2496
FT NONHELICAL REGION (NC3).
FT DOMAIN 2451 2746
FT TRIPLE-HELICAL REGION (COL2)
FT DOMAIN 2747 2898
FT WITH 1 IMPERFECTION.
FT DOMAIN 2899 2941
FT NONHELICAL REGION (NC2).
FT DOMAIN 2942 3044
FT TRIPLE-HELICAL REGION (COL1)
FT WITH 2 IMPERFECTIONS.
FT DOMAIN 3045 3063
FT NONHELICAL REGION (NC1).
FT DOMAIN 798 798
FT TO CHONDROITIN SULFATE (POTENTIAL).
FT BINDING 889 889
FT TO CHONDROITIN SULFATE (POTENTIAL).
FT BINDING 981 981
FT TO CHONDROITIN SULFATE (POTENTIAL).
FT SITE 862 864
FT CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2779 2781
FT CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2895 2897
FT CELL ATTACHMENT SITE (POTENTIAL).
FT MOD_RES 2944 2944
FT HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2947 2947
FT HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2950 2950

```

```
FT MOD_RES 2959 2959 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2965 2965 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2968 2968 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2971 2971 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2983 2983 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3000 3000 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3003 3003 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3014 3014 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3023 3023 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3026 3026 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3029 3029 HYDROXYLATION (BY SIMILARITY).
FT CARBOHYD 700 700 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1763 1763 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2206 2206 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2528 2528 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2679 2679 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 25 1188 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 3063 AA; 333189 MW; 75FEA78FA8E48293 CRC64;

Query Match 6.88; Score 131; DB 1; Length 3063;
Best Local Similarity 22.18; Pred. No. 0.032;
Matches 74; Conservative 71; Mismatches 116; Indels 74; Gaps 18;

Qy 44 DLYFTLDKSGSV-LHHMNEIYFVEQLAKHFI--SPQLRMSFIVFSTRGTTLMKLT--D 98
Db 140 DLVELVDGWSVGRNFKYILDFDAALVSADFGEKTRGVGVQYSSDTRTEFNLNQYYQ 199
Qy 99 REQIRQGLEELQKV-LPGDGYTHMEGFRASEQIYEN---RQGYRTASVTIATDGEIH 154
Db 200 RDEL---LAAIKKIPYKGGTMTGDAIDYLVKNTFTESAGARVGPVKVAILII--TDGK 254
Qy 155 EDLFFYSEREANRSDLGAIYVCVCKDFNETQLARIAD--SKOHVFPVNDGFOALQII 212
Db 255 DEV-----EIPARELNRNMGVEVFSLGKRAADAKELQIASTPSLNHVFNVAN-FDAI 309
Qy 213 HSLKKSCI-----ETLAAE-----PSTICAGESFQVVRGN-----GFR--- 247
Db 310 NEITSQVCGVDEQLGELVSGVEEVPSPNLIAEVSRYKVLNWNPSPTVGYKVILT 369
Qy 248 -----HARNVRVLCSFKINDSVTLNEKPFVEDTYLPCAPILKEVGKKAALQVSM 299
Db 370 PMTAGSRQHALLSVG-----PQTTLSVRDLSDADTEY-----QISVSAMKGMTS 412
Qy 300 NDGLSFI-SSSVIITTHCSGSIILAIALLILFL 333
Db 413 SEPISIMEKTOPMKVQVCECSRG--VDIKADIVFLV 445

RESULT 12
CAIE_CHICK
ID CAIE_CHICK STANDARD; PRT: 1888 AA.
AC P32018;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(XIV) chain precursor (Undulin).
GN COL14AL.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=93185668; PubMed=8444186;
RA Waelchli C., Trueb J., Kessler B., Winterhalter K.H., Trueb B.;
RT "Complete primary structure of chicken collagen XIV.";
RL Eur. J. Biochem. 212:483-490(1993).
RN [2]
RP SEQUENCE OF 402-1549 FROM N.A.
RX MEDLINE=92339443; PubMed=1339349;
```

```
RA Trueb J., Trueb B.;
RT "Type XIV collagen is a variant of undulin.";
RL Eur. J. Biochem. 207:549-557(1992).
RN [3]
RX STRUCTURE BY NMR OF 1852-1885.
RP MEDLINE=99280705; PubMed=10350466;
RA Montserret R., Aubert-Foucher E., McLeish M.J., Hill J.M., Ficheux D.,
Jaguinod M., van der Rest M., Deleage G., Penin F.;
RT "Structural analysis of the heparin-binding site of the NCI domain of
collagen XIV by CD and NMR.";
RL Biochemistry 38:6479-6488(1999).
CC -1- FUNCTION: AN ADHESIVE ROLE BY INTEGRATING COLLAGEN BUNDLES. IT IS
PROBABLY ASSOCIATED WITH THE SURFACE OF INTERSTITIAL COLLAGEN
FIBRILS VIA COLL. THE COL2 DOMAIN MAY THEN SERVE AS A RIGID ARM
WHICH STICKS OUT FROM THE FIBRIL AND PROTRUDES THE LARGE N-
TERMINAL GLOBULAR DOMAIN INTO THE EXTRACELLULAR SPACE, WHERE IT
MIGHT INTERACT WITH OTHER MATRIX MOLECULES OR CELL SURFACE
RECEPTORS.
CC -1- SUBUNIT: HOMOTRIMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: Extracellular matrix.
CC -1- TISSUE SPECIFICITY: WIDE TISSUE DISTRIBUTION; HIGH PRESENCE IN
DENSE CONNECTIVE TISSUE IN SKELETAL MUSCLE.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- PTM: LYSINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
UNIT ARE HYDROXYLATED IN ALL CASES AND BIND CARBOHYDRATE.
CC -1- PTM: MAY CONTAIN CYSTEINE RESIDUES INVOLVED IN INTER- AND
INTRAMOLECULAR DISULFIDE BONDING.
CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
INTERRUPTED HELICES (FACIT) FAMILY.
CC -1- SIMILARITY: CONTAINS 2 VWFA DOMAINS.
CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
EMBL; X70793; CAA50064.1; -
EMBL; X70792; CAA50063.1; -
EMBL; X66138; CAA46928.1; -
PIR; S22916; S22916
PDB; 1B9P; 25-FEB-99.
PDB; 1B9Q; 25-FEB-99.
InterPro; IPR000087; Collagen.
InterPro; IPR003961; FN_III.
InterPro; IPR003962; FNIII_repeat.
InterPro; IPR003129; TSPN.
InterPro; IPR002035; VWFA.
Pfam; PF01391; Collagen; 4.
Pfam; PF00041; fn3; 8.
Pfam; PF02210; TSPN; 1.
Pfam; PF00092; vwa; 2.
PRINTS; PR00014; FNTPPEIII.
PRINTS; PR00453; VWFADOMAIN.
SMART; SM00060; FN3; 7.
SMART; SM00210; TSPN; 1.
SMART; SM00327; VWA; 2.
PROSITE; PS50234; VWFA; 2.
Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
Cell adhesion; Collagen; Glycoprotein; Signal; 3D-structure.
KW SIGNAL
FT CHAIN 1 28
FT DOMAIN 29 1888
FT DOMAIN 158 330
FT DOMAIN 441 529
FT DOMAIN 531 622
FT DOMAIN 625 728
FT DOMAIN 739 821
FT DOMAIN 829 922
FT DOMAIN 952 1010
```

FT DOMAIN 1042 1215 VWFA 2.  
FT DOMAIN 1227 1468 NONHELICAL REGION (NC4).  
FT DOMAIN 1469 1620 TRIPLE-HELICAL REGION 1.  
FT DOMAIN 1664 1786 TRIPLE-HELICAL REGION 2.  
FT DOMAIN 721 733 POLY-THR.  
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1398 1398 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT SITE 1489 1491 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 1617 1619 CELL ATTACHMENT SITE (POTENTIAL).  
SQ SEQUENCE 1888 AA; 202666 MW; 3991589F46DD873 CRC64;

Query Match 6.8%; Score 130; DB 1; Length 1888;  
Best Local Similarity 25.0%; Pred. No. 0.021;  
Matches 62; Conservative 44; Mismatches 116; Indels 26; Gaps 12;

QY 44 DLVFLDKSGV-LHWNNEIYFVEQL--AHKFISPO-LRMSFIVFSTRTTLMKLT--E 97  
Db 1042 DLVFLDGSWSIGDNNKLIISFLYSTVGALDKIGPDGTQVAIIQFSDDPRTFELKNAYK 1101  
QY 98 DREGIROGLELQKVLPGGDTYMEGFERASEQIYYENROGYTA--SVIALTDLGELHE 155  
Db 1102 TKETLLEAIQI--AYKGGNTKTKAIKHARE-VLFTGEAGMRKGIPKVLVIIDGRSQD 1158  
QY 156 DLFFYSEREANRSDL-GAIYVCYGVDFNETOLARIAD--SKDHVPVNDGFOALQGI 212  
Db 1159 DV-----NKVSREMOLDGFSFAIGVADADYSELVNIIGSKPSERHVFVDD-PDAFTKIE 1212  
QY 213 HSIILKCIETILAEPTICAGESFQVVRGNGFRHARNVDRLVLCSPKINDSVTLNKKPF 272  
Db 1213 DELITFVCETASATCPLVKDGDKLA-----GFKMMEMFGLVEKEFSAIDGVSMERGTF 1266  
QY 273 SVEDTYLL 280  
Db 1267 NVYPCYRL 1274

RESULT 13  
ID YNX3\_CABEL STANDARD; PRT: 3051 AA.  
AC P34576;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein T20G5.3 in chromosome III (Fragment).  
GN T20G5.3  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Berks M., Smith A.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: CONTAINS 45 EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 2 SEA DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL; 230423; CAA83007.1; -.  
DR PIR; S42373; S42373.  
DR HSP; P02468; I1LE.  
DR WormPep; T20G5.3; CE00478.  
DR InterPro; IPR000152; Asx\_hydroxyl.  
DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR001881; EGF\_Ca.  
DR InterPro; IPR000082; SEA.  
DR InterPro; IPR002035; VWFA.  
DR Pfam; PF00008; EGF; 30.  
DR Pfam; PF01390; SEA; 2.  
DR Pfam; PF00092; vwa; 1.  
DR SMART; SM00179; EGF\_CA; 1.  
DR SMART; SM00001; EGF\_Like; 38.  
DR SMART; SM00200; SEA; 2.  
DR SMART; SM00327; VWFA; 1.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 28.  
DR PROSITE; PS00022; EGF\_1; 1.  
DR PROSITE; PS01186; EGF\_2; 5.  
DR PROSITE; PS01187; EGF\_CA; 1.  
DR PROSITE; PS00024; SEA; 2.  
DR PROSITE; PS00334; VWFA; 1.  
KW Hypothetical protein; EGF-like domain; Repeat; Transmembrane.  
FT NON\_TER 1  
FT DOMAIN <1 2701 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 2702 2722 POTENTIAL.  
FT DOMAIN 2723 3051 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN <1 44 EGF-LIKE 1 (PARTIAL).  
FT DOMAIN 46 94 EGF-LIKE 2.  
FT DOMAIN 100 144 EGF-LIKE 3.  
FT DOMAIN 145 192 EGF-LIKE 4.  
FT DOMAIN 194 245 EGF-LIKE 5.  
FT DOMAIN 247 296 EGF-LIKE 6.  
FT DOMAIN 313 354 EGF-LIKE 7.  
FT DOMAIN 355 402 EGF-LIKE 8.  
FT DOMAIN 404 452 EGF-LIKE 9.  
FT DOMAIN 454 503 EGF-LIKE 10.  
FT DOMAIN 514 690 VWFA.  
FT DOMAIN 705 750 EGF-LIKE 11.  
FT DOMAIN 750 794 EGF-LIKE 12.  
FT DOMAIN 805 846 EGF-LIKE 13.  
FT DOMAIN 847 892 EGF-LIKE 14.  
FT DOMAIN 892 940 EGF-LIKE 15.  
FT DOMAIN 942 990 EGF-LIKE 16.  
FT DOMAIN 992 1039 EGF-LIKE 17.  
FT DOMAIN 1043 1091 EGF-LIKE 18.  
FT DOMAIN 1093 1144 EGF-LIKE 19.  
FT DOMAIN 1146 1195 EGF-LIKE 20.  
FT DOMAIN 1197 1245 EGF-LIKE 21.  
FT DOMAIN 1247 1295 EGF-LIKE 22.  
FT DOMAIN 1298 1346 EGF-LIKE 23.  
FT DOMAIN 1352 1396 EGF-LIKE 24.  
FT DOMAIN 1397 1444 EGF-LIKE 25.  
FT DOMAIN 1446 1492 EGF-LIKE 26.  
FT DOMAIN 1494 1542 EGF-LIKE 27.  
FT DOMAIN 1544 1592 EGF-LIKE 28.  
FT DOMAIN 1594 1642 EGF-LIKE 29.  
FT DOMAIN 1644 1692 EGF-LIKE 30.  
FT DOMAIN 1693 1739 EGF-LIKE 31.  
FT DOMAIN 1740 1788 EGF-LIKE 32.  
FT DOMAIN 1787 1847 EGF-LIKE 33.  
FT DOMAIN 1849 1900 EGF-LIKE 34.  
FT DOMAIN 1902 1950 EGF-LIKE 35.  
FT DOMAIN 1952 1998 EGF-LIKE 36.  
FT DOMAIN 2000 2047 EGF-LIKE 37.  
FT DOMAIN 2047 2095 EGF-LIKE 38.  
FT DOMAIN 2117 2156 EGF-LIKE 39.  
FT DOMAIN 2157 2281 SEA 1.  
FT DOMAIN 2293 2332 EGF-LIKE 40.  
FT DOMAIN 2333 2456 SEA 2.  
FT DOMAIN 2460 2504 EGF-LIKE 41.  
FT DOMAIN 2508 2556 EGF-LIKE 42.  
FT DOMAIN 2556 2608 EGF-LIKE 43.  
FT DOMAIN 2612 2657 EGF-LIKE 44, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 2657 2693 EGF-LIKE 45.  
FT DISULFID 2 13 BY SIMILARITY.  
FT DISULFID 7 22 BY SIMILARITY.  
FT DISULFID 24 43 BY SIMILARITY.  
FT DISULFID 50 63 BY SIMILARITY.





Query Match	6.3%	Score 120.5;	DB 1;	Length 496;
Best Local Similarity	25.3%	Pred. NO. 0.023;		
Matches	60;	Conservative 37;	Matches 95;	Indels 45; Gaps 13;
QY	24	ICAGGGRREDGGPACVGGFDLYFLDKSSVL-HHNEIYYFVEOLAHKF-ISPQLRMS	81	
	:	: :	:	:
	:	: :	:	:
	:	: :	:	:
	:	: :	:	:
Db	264	VCSGGGS-----SATDLVDLGDGKSVRPENFELVKKEISQIVDTLDVSKLAQV	314	
	:	: :	:	:
	:	: :	:	:
	:	: :	:	:
	:	: :	:	:
QY	82	FIV--FSTRGTTLMLKLTEDREQRLQGLELQKLPGDGYMHGEFRASEQIY-----Y	133	
	:	: :	:	:
	:	: :	:	:
	:	: :	:	:
	:	: :	:	:
Db	315	GLVQYSSVVQRQEPPLGRHTTKDKAAVRNM-----SYMEKGMTGAALKYLIDNSF	366	
	:	: :	:	:
	:	: :	:	:
	:	: :	:	:
	:	: :	:	:
QY	134	ENRQGYRTAS--VIALTGDGELHEDLFYSREANRSRDLAGIYCVGVKDFNETOLARI	191	
	:	: :	:	:
	:	: :	:	:
	:	: :	:	:
	:	: :	:	:
Db	367	TVSSGARPGAQKVGIPTDGSRSD----YINDAKKAADLGFKMFVGAVGNAVEDEREI	422	
	:	: :	:	:
	:	: :	:	:
	:	: :	:	:
	:	: :	:	:
QY	192	ADS--KDHVPFYNDGFQALOGIIHSILKSCSIEILAAEPESTICAGES---FQVVVRG	243	
	:	: :	:	:
	:	: :	:	:
	:	: :	:	:
	:	: :	:	:

RESULT 15		
CAIC_NOTVI	STANDARD;	PRT; 929 AA.
ID CAIC_NOTVI		
AC Q91145;		
DT 15-JUL-1998	(Rel. 36, Created)	
DT 15-JUL-1998	(Rel. 36, Last sequence update)	
DT 01-MAR-2002	(Rel. 41, Last annotation update)	
DE Collagen alpha 1(XI) chain (Fragment).		
OS Notophthalmus viridescens (Eastern newt) (Triturus viridescens).		
DE Nototheniidae		
DE Teleostei		
DE Euteleostomi		
DE Vertebrata		

OC Amphibia; Batrachia; Caudata; Salamandroidae; Salamandridae;  
OC Notophthalmus;  
OC NCBI\_TaxID=8316;  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=95246925; PubMed=7729585;  
RX Wei Y., Yang E.V., Klatt K.P., Tassava R.A.;  
RT "Monoclonal antibody MT2 identifies the urodele alpha 1 chain of type  
FA XII collagen, a developmentally regulated extracellular matrix  
RT protein in regenerating newt limbs.";  
RL Dev. Biol. 168:503-513(1995).  
CC -!- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-  
CC CONTAINING FIBRILS, THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE  
CC SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE  
CC LOCALIZED IN THE PERIFIBRILLAR MATRIX (BY SIMILARITY). COULD PLAY  
CC A DEVELOPMENTAL ROLE IN REGENERATION.  
CC -!- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 kDa OF  
CC NONTRIPLE-HELICAL SEQUENCES (BY SIMILARITY).  
CC -!- DEVELOPMENTAL STAGE: EXPRESSION STARTS AT 3 DAYS AFTER AMPUTATION  
CC IN CELLS OF THE BASAL LAYER OF THE WOUND EPITHELIUM. AT DAY 10,  
CC EXPRESSION IS FOUND IN BOTH THE BASAL WOUND EPITHELIAL CELLS AND  
CC THE DISTAL MESenchyme CELLS. AT MID-BUD AND LATE-BUD BLASTOMA  
CC STAGES, WOUND EPITHELIUM EXPRESSION HAS DECREASED, WHEREAS THE  
CC MESenchyme REMAINS SPONGY ACTIVE IN TRANSCRIPTION AND SHOWS A  
CC TENDENCY TOWARD DISTAL REGIONALIZATION. CONDENSING CARTILAGE SHOWS  
CC NO SIGNAL. FINALLY, AT THE LATE DIGIT STAGE, EXPRESSION BECOMES  
CC LARGELY RESTRICTED TO THE PERICHONDRUM.  
CC -!- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT  
CC EACH END (BY SIMILARITY).  
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY  
CC SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH  
CC INTERRUPTED HELICES (FACIT) FAMILY.  
CC -!- SIMILARITY: CONTAINS 2 VWFA DOMAINS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: U19494; AAA80217.1; \*.  
DR HSP: P02751; LFNA.  
DR InterPro: IPR003961; FN.III.  
DR InterPro: IPR002035; VWFA.  
DR Pfam: PF00041; fn3; 7.  
DR Pfam: PF00092; vwa; 1.  
DR SMART: SM00060; FN3; 5.  
DR SMART: SM00327; VWA; 1.  
DR PROSITE: PS50234; VWFA; 2.  
DR Extracellular matrix; Connective tissue; Repeat; Cell adhesion;  
KW Collagen; Glycoprotein.  
FT NON\_TER 1  
FT DOMAIN <1 49  
FT DOMAIN 63 154  
FT DOMAIN 155 245  
FT DOMAIN 246 338  
FT DOMAIN 339 432  
FT DOMAIN 433 519  
FT DOMAIN 520 612  
FT DOMAIN 633 805  
FT DOMAIN 818 907  
FT DOMAIN 908 >929  
FT BINDING 231 231  
FT BINDING 324 324  
FT BINDING 415 415  
FT CARBOHYD 98 98  
FT NON\_TER 929 929  
SQ SEQUENCE 929 AA; 101647 MW; AESD7485254FD954 CRC64;

Search completed: August 9, 2002, 10:46:58  
Job time: 821 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 8, 2002, 20:55:43 ; Search time 4666.15 seconds  
(without alignments)  
6341.443 Million cell updates/sec

Title: US-09-970-076-1

Perfect score: 1414

Sequence: 1 aggaccgcgaggaaggcc.....aaaaaaaaaaaaaaaaaaaa 1414

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_on.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
-----							

1	1414	100.0	1414	9	AF421380	Homo sapi
2	1193.8	84.4	5540	9	AF279145	Homo sapi
3	1052.4	74.4	2112	9	BC012074	Homo sapi
4	890	62.9	1436	9	AK001463	Homo sapi
5	853.4	60.4	5220	10	AF378762	Mus muscu
6	542.8	38.4	4007	9	AK025429	AK025429 Homo sapi
7	255.8	18.1	2126	9	AK057316	BC012475 Homo sapi
8	236	16.7	1297	9	BC012475	BC012475 Homo sapi
9	232.2	16.4	2026	9	AK055636	AK055636 Homo sapi
10	207.8	14.7	1343	9	AY040326	AY040326 Homo sapi
11	194.4	13.7	81017	2	AC025010	AC025010 Homo sapi
12	194.4	13.7	81017	2	AC025010	AC025010 Homo sapi
13	113	8.0	1189	9	AK002160	AK002160 Homo sapi
14	92.8	6.6	2557	10	BC003908	BC003908 Mus muscu
15	62	4.4	162448	30	AC053530	AC053530 Homo sapi
16	61.2	4.3	1512	10	BC010278	BC010278 Mus muscu
17	61	4.3	8712	6	AX344694	AX344694 Sequence
18	60.2	4.3	52359	2	AC010772	AC010772 Homo sapi
19	60	4.2	67676	2	AC102488	AC102488 Mus muscu
20	59.4	4.2	6070	6	AX281468	AX281468 Sequence
21	59.4	4.2	6070	6	AX346581	AX346581 Sequence
22	59.4	4.2	6070	6	AX348804	AX348804 Sequence
23	59.4	4.2	6668	6	AX346598	AX346598 Sequence
24	59.4	4.2	13712	6	AX346433	AX346433 Sequence
25	59.2	4.2	9219	6	AX251565	AX251565 Sequence
26	59.2	4.2	9219	6	AX347328	AX347328 Sequence
27	59	4.2	349980	6	AX344558	AX344558 Sequence
28	58.4	4.1	74412	2	AC020767	AC020767 Homo sapi
29	58.2	4.1	611	6	AX187051	AX187051 Sequence
30	58	4.1	6283	6	AX345737	AX345737 Sequence
31	58	4.1	6283	6	AX348379	AX348379 Sequence
32	58	4.1	6944	6	AX347467	AX347467 Sequence
33	58	4.1	6944	6	AX349188	AX349188 Sequence
34	57.8	4.1	7319	6	AX348494	AX348494 Sequence
35	57.8	4.1	7736	3	AF310889	AF310889 Dictyoste
36	57.6	4.1	7076	6	AX347047	AX347047 Sequence
37	57.6	4.1	11015	6	AX280003	AX280003 Sequence
38	57.6	4.1	11015	6	AX281190	AX281190 Sequence
39	57.6	4.1	11015	6	AX356459	AX356459 Sequence
40	57.6	4.1	349980	6	AX344560	AX344560 Sequence
41	57.2	4.0	70816	2	AC101610	AC101610 Mus muscu
42	57.2	4.0	349980	6	AX344556	AX344556 Sequence
43	57	4.0	2379	3	DD31	X54452 D.discoldeu
44	57	4.0	89128	2	AP001826	AP001826 Homo sapi
45	57	4.0	126038	2	AC016068	AC016068 Homo sapi

#### ALIGNMENTS

#### RESULT 1

AF421380	AF421380	1414 bp	mrna	linear	PRI 13-NOV-2001
LOCUS	Homo sapiens	anthrax toxin receptor mRNA	complete cds.		
DEFINITION	Nature 414 (6860), 225-229 (2001)				
ACCESSION	AF421380				
VERSION	AF421380.1	GI:16566412			
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 1414)				
AUTHORS	Bradley, K.A., Mogridge, J., Mourez, M., Collier, R.J. and Young, J.A.T.				
TITLE	Identification of the cellular receptor for anthrax toxin				
JOURNAL	Nature 414 (6860), 225-229 (2001)				
MEDLINE	21557240				
PUBMED	11700562				
REMARK	http://www.nature.com				
REFERENCE	2 (bases 1 to 1414)				
AUTHORS	Bradley, K.A., Mogridge, J., Mourez, M., Collier, R.J. and Young, J.A.T.				
TITLE	Direct Submission				
JOURNAL	Submitted (19-SEP-2001) Department of Oncology, University of Wisconsin-Madison, 1400 University Ave., Madison, WI 53706, USA				

FEATURES		Location/Qualifiers	
source		1..1414	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/cell_line="HeLa"	
5'UTR		1..103	
CDS		104..1210	
		/note="ATR"	
		/codon_start=1	
		/product="anthrax toxin receptor"	
		/protein_id="AAL26496.1"	
		/db_xref="GI:16566413"	
		/translation="NATARRALIGFQWLSLATLVLCAGCGRRDGGPACVGGDF LVFLDKSGSVLHNNIYFVEQLAHKFIQPOLRMSEFVSTRTGTLMLKTEDREQI RGLERLQVLPGDGTWHEGFEDASEQIYYENROGYRTASVIIALTDEGHEDLFFY SEREANKRDLGAIYCVGVKDFENETOLARIADSKHVPVNDGFQALQGIHSILKK SCIEILAAPESTICAGESQVVVRNGRPHARNVDRVLCSPKINDSVTLNKPFSVED TYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVLIYTHCSDGSLAIALLLFLFL ALALLWFWPLCCTVIKEVPPPPAPESEENKIK"	
3'UTR		1211..1414	
BASE COUNT	394 a 344 c 344 g 332 t		
ORIGIN			
Query Match 100.0%; Score 1414; DB 9; Length 1414;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 1414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	aggaccgcgagggaagggccgcgagatggcgctccctgagggctggtggcgagttcgcg 60	
DB	1	AGGACCCGCGAGGAAGGGCCCGGATGGCGCTCCCTGAGGGTCGTGGGAGTTCGCGG 60	
QY	61	agcgtgggaagagcgagccctgctctcccgggctgcggcgccatgcccagcgcgagcg 120	
DB	61	ACCGTGGGAGAGCGGACCCCTGCTCTCCCGGGCTGCGGCCATGGCCACGGCGAGCG 120	
QY	121	gagagccctcgcatcggtctccagtgcctctctttggccactctggtgctcatctgcgc 180	
DB	121	GAGAGCCCTCGCATCGGCTTCCATGGCTCTCTTTGGCCACTCTGGTGCCTCATCTGCGC 180	
QY	181	cgggcaaggggagcagggagatgggggtccagcctgctacggcgagttgacctgta 240	
DB	181	CGGGCAAGGGGACGAGGAGATGGGGTCCAGCCTGCTACGGCGGATTTGACCTGTA 240	
QY	241	cttcaatttggacaatacaggaagtgtctgcaccactggaatgaaatctattacttgt 300	
DB	241	CTTCAATTTGGACAAATCAGGAAGTGTCTGCACCACCTGGAATGAATCTATTACTTTGT 300	
QY	301	ggacagttggctcacaattcatcagccacagttgagaaatgtcctttattgtttctc 360	
DB	301	GGAACAGTTGGCTCACAATTCATCAGCCACAGTTGAGAATGCTCTTTATTGTTTCTC 360	
QY	361	cacccgagggaacaccttaataaactgacagaagacagacaataccgtcaaggcct 420	
DB	361	CACCCGAGGAACAACCTTAATGAACCTGACAGAAGACAGACACAATAATCCGTCAAGCCCT 420	
QY	421	agaagaactccagaagttctgcgagggagagacacttacatgcatgaaggattgaaag 480	
DB	421	AGAAGAATCCAGAAAGTTCTGCCAGGAGAGACACTTACATGCATGAAGGATTTGAAAG 480	
QY	481	ggccagtgcagatattattataaacaagaggtacagacagccagcgctcatcat 540	
DB	481	GGCCAGTGAGCAGATTTATTATGAACACAGCAAGGGTACAGGACAGCCAGCGTCATCAT 540	
QY	541	tgttttaactgatgagaactccatgaatctctttttatttcagagagggagagctaa 600	
DB	541	TGCTTTGACTGATGGAACCTCCATGAATCTCTTTTCTATTTCAGAGAGGGAGGCTAA 600	
QY	601	taggtctcagatcttggtgcaattgtttactgtgttggtgtaaaattcaatgaagac 660	
DB	601	TAGGTCTCGATCTTGTCGAATTGTTTACTGTGTGGTGTGAAGATTTCATATGAGAC 660	
QY	661	acagctggcccgattgcggacagtaagatcatgtgtttccctgtaagcggtttca 720	

661	ACAGCTGCCCGGATTCGGACAGTAAGGATCATGTGTTTCCCGTGAATGACGGCTTTCA	720
721	ggctctcaagggaatcatccactcaattttgaagaagtcctgcataaattctagcagc	780
721	GGCTCTGCAAGGCATCATCCACTCAATTTTGAAGAAGTCTGTCATCGAATTTCTAGCAGC	780
781	tgaaccatccaccatgtgcagagagtcatttcaagttgtcgtgagagaaacggcctt	840
781	TGAACCATCCACATATGTGCAGGAGAGTCATTTCAAGTTGTCTGTGAGAGAAACGGCTT	840
841	ccgacatgcggcgcaacgtggacaggggtcctctgcaggttcaaatcaatgactcgtcac	900
841	CCGACATGCCCGCAACGTGGACAGGGTCTCTGCAGCTTCAAGATCAATGATCGGTTCAC	900
901	actcaatgagaagccctttctgtggaagacactatttactgtgtccagcgccatcttt	960
901	ACTCAATGAGAAGCCCTTTTCTGTGGAAGACACTATTATTACTGTGTCCAGCGCTATCTT	960
961	aaagaagttggcatgaaagctgcactccaggtcagcatgaacatgagcctctctttat	1020
961	AAAGAAGTTGGCATGAAGCTGCACCTCCAGTCCAGATGAACGATGGCCTCTCTTTAT	1020
1021	ctccagttctcatcatcaccacacacactgttctgaacgttccatcctcgccatcgc	1080
1021	CTCCAGTTCTCATCATCACCACCACACACTGTCTTCTGACGGTTCATCTCGGCATCGC	1080
1081	cctgctgacccctgttctcctcctagccctgctcctcctctcctctcctcctctctg	1140
1081	CCTGCTGATCCCTGTTCTCTGCTCTAGCCCTGGCTCTCTCTGTTGTTCTGCGCCCTCTG	1140
1141	ctgcactgtgattatacaaggaggtccctccaccctgcgagggagtgaggaataaa	1200
1141	CTGCACTGTGATTATCAAGGAGTCCCTCCACCCCTGCCGAGGAGTGAGGAAAAATAA	1200
1201	aataaatacaagaagaagaagaataatccacagacaagataacacacacagc	1260
1201	AATAAATAACAGAGAGAAAGAAAGAAATCCACAGAAACAGATAACCTTAACACAGC	1260
1261	ccgtgcaacgtattttatacaatgctcgtgaaatcagtctcaatcctagacagctttt	1320
1261	CCGTGCAACGTATTTATACAATGCTCTGAAATCATAGTCTCAATCTAGACAGTCTTTT	1320
1321	ccctagtctcctgtattcaatccaggtgtctaaatcattcaataaatagctatatgaat	1380
1321	CCCTAGTTCCTGTATTCAATCCCAAGTCTTAACATTCATAATAATAGCTATATGAAT	1380
1381	caaaaaaaaaaaaaaaaaaaaaaaaaaaaaa	1414
1381	CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	1414
RESULT 2		
AF279145	AF279145	5540 bp mRNA linear PRI 09-MAY-2001
LOCUS	Homo sapiens tumor endothelial marker 8 precursor (TEM8), mRNA,	
DEFINITION	complete cds.	
ACCESSION	AF279145	
VERSION	AF279145.2	GI:14017380
KEYWORDS		
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 5540)	
AUTHORS	St. Croix,B., Rago,C., Velculescu,V., Traverso,G., Romans,K.E., Montgomery,E., Lal,A., Riggins,G.J., Lengauer,C., Vogelstein,B. and Kinzler,K.W.	
TITLE	Genes expressed in human tumor endothelium	
JOURNAL	Science 289 (5482), 1197-1202 (2000)	
MEDLINE	20407466	
PUBMED	10947988	
REFERENCE	2 (bases 1 to 5540)	

AUTHORS St. Croix,B., Vogelstein,B. and Kinzler,K.W.  
TITLE Direct Submission  
JOURNAL Hopkins University, 1650 Orleans Street, Baltimore, MD 21231, USA  
REFERENCE 3 (bases 1 to 5540)  
AUTHORS St. Croix,B., Vogelstein,B. and Kinzler,K.W.  
JOURNAL Direct Submission  
REMARK Hopkins University, 1650 Orleans Street, Baltimore, MD 21231, USA  
COMMENT Sequence update by submitter version replaced gi:9857405.  
FEATURES  
    Location/Qualifiers  
        1..5540  
            /organism="Homo sapiens"  
            /db\_xref="taxon:9606"  
            /chromosome="4"  
        1..5540  
            /gene="TEM8"  
        144..1838  
            /gene="TEM8"  
        /codon\_start=1  
        /product="tumor endothelial marker 8 precursor"  
        /protein\_id="AAK2094.1"  
        /db\_xref="GI:14017381"  
        /translation="MATAERRALGIFQWLSLATLVLCAGQGRREDGGPACYYGGFD  
LYFLDKSGVLHWNIEYFVEQLAHKFIQPOLRMSFVFEVSTRTGLMLKLTEDREQI  
ROGLEELQVLPGDGYMHGFEFRASEIOIYVENRGVYTSVIALTDGELHEDFEY  
SERANRSRLGAIYVCGVDFNETQARIADSKDQVDFVNDGFQALQGLIHLKK  
SYIELAAPETICAGSFQVVVNGRNGFARNVDVLCSEKINDSVTLNKPFSVED  
TYLCLAPILKEVKMKAALQVSMNDGLSFSSVITTHCSGSLIALILLILFLLL  
ALALLWFWLCCVTIIEVPPPAESEEEDDGLPKKWPPTDASYGGGRVGGIK  
RMEVWKEGKSTEGAKLEKARVKMPEOEYEFPEPNLNMRNRPSSPRKVSPI  
KGKLDLWLLKRYGDRVSNRPOGDTGRICNFRVKNOPAKYPLNNAVHTSSPPP  
APIYTPPPAPHCPPSPSAPTPIPPSPSLPPPPQAPPPNRPSPRPPSPSV"  
        144..224  
            /gene="TEM8"  
        225..1835  
            /gene="TEM8"  
        /product="tumor endothelial marker 8"  
        1104..1172  
            /gene="TEM8"  
        /note="Region: predicted transmembrane domain"  
BASE COUNT 1601 a 1306 c 1187 g 1446 t  
ORIGIN  
  
Query Match 84.4%; Score 1193.8; DB 9; Length 5540;  
Best Local Similarity 99.8%; Pred. No. 2.9e-306;  
Matches 1195; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 aggaaccgagagagggccgagatggcgctccctgagggctgctgaggttcgaggttcgcg 60  
|||||  
DB 41 AGACCCGGAGAGAGGGCCCGGATGGCGGCTCCTGTAGAGGTGCTGGCGAGTTCGCGG 100  
|||||  
QY 61 agctgggaagagcgagccctgctctccccgggctgctggcgccatggccacggcgagcg 120  
|||||  
DB 101 AGCGTGGGAAGGAGCGACCCCTGCTCTCCCGGGCTGCGGGCCATGTCGCCGCGAGCG 160  
|||||  
QY 121 gagagccctcgagcagcttcctccagtgctctctcttttggccactctgggtctatctgcg 180  
|||||  
DB 161 GAGAGCCCTCGGCATCGGCTTCACATGGCTCTCTTTGGCCACTCTGGTGTCTATCTGCGC 220  
|||||  
QY 181 cgggcaaggggacagagagagtgagggtccagcctgctacgctgctacgctgagctgta 240  
|||||  
DB 221 CGGGCAAGGGGACGAGGAGATGGGGGTCCAGCCTGCTACGGCGGATTTGACCTGTA 280  
|||||  
QY 241 cttcattttggacaaatcagggaagtgtgtgcaccactggaatgaaatctattcttctgt 300  
|||||  
DB 281 CTTTCATTTGGCAATCAGGAAGTGTGCTGCACCACCTGGAATGAAATCTATTACTTTGT 340  
|||||  
QY 301 ggaacagttggtcacaattcatcagcccaacagttgagaatgctctttattgtttctc 360  
|||||  
DB 341 GGAACAGTTGGTCAAAATTCAGGCCACAGTTGAGAAATGTCTTTATTGTTTCTC 400  
|||||

QY 361 caccgaggaacaaaccttaataaactgacagaacacagagacaaaatccgtcaaggcct 420  
|||||  
DB 401 CACCCGAGGAACAACCTTAATGAACACTGACAGAAGACAGAGACAAATCCGTCAGGCCT 460  
|||||  
QY 421 agaagaactccagaagaagttctccaggaggagacacttacatgcatgaaggatttgaaag 480  
|||||  
DB 461 AGAAGAAGTCCAGAAGTCTGCCAGGAGGAGACACTTACATGTCATGAAGGATTTGAAG 520  
|||||  
QY 481 ggccaagtgaagatttattatgaaaaacagacaagaagggtacagagacagccagctcatc 540  
|||||  
DB 521 GGCCAGTGAGCAGATTATTATGAACACAGACAAGGGTACAGGACAGCCAGCGTCATCAT 580  
|||||  
QY 541 tgccttgactgatggagaactccatgaagatctcttttctattcagagagggaggtctaa 600  
|||||  
DB 581 TGCCTTGACTGATGGAGAACTCCATGAAGATCTCTTTTCTATTTCAGAGAGGGAGCTAA 640  
|||||  
QY 601 tagtctcgagatcttggtgcaattgtttactgtgtgtgtgtgaaagatttcaatgaagac 660  
|||||  
DB 641 TAGGTCCTGAGATCTTGGTGAATTTGTTTACTGTGTGTTGTTGTTGTTGTTGTTGTTG 700  
|||||  
QY 661 acagtggtgcccagattggcgacagtaagatcatgtgtttcccggtgaatgacggtttca 720  
|||||  
DB 701 ACAGCTGGCCCCGATGTCGGACAGTAAGGATCATGTGTTTCCGCTGAATGACGGCTTCA 760  
|||||  
QY 721 ggctctgaaggcctacactcaactcaatttgaagtcctctgcacgcgaattcttagcagc 780  
|||||  
DB 761 GGCTCTGCAGGACATCATCCACTCAATTTGAAGAAGTCTCGATCGAAATTTAGCAGC 820  
|||||  
QY 781 tgaacattccaccatgatgtgcaggagagtcatttcaagttgtgtgagaggaagagcgtt 840  
|||||  
DB 821 TGAACATCCACCATATGTGCAGGAGAGTCAATTTCAAGTTGTGCTGAGAGAAACGGCTT 880  
|||||  
QY 841 ccgacatgcccgaacgctgagagggctctctgcagcttcaagatcaatgaactcggtcac 900  
|||||  
DB 881 CCGACATGCCCAACGCTGGACAGGGTCCCTCTGCAGCTTCAAGATCAATGACTCGGTCA 940  
|||||  
QY 901 actcaatgagaagccctttctgtggaagacactatttactgtgtccagcgccctatctt 960  
|||||  
DB 941 ACTCAATGAGAAGCCCTTTCTGTGGAAGTACTTATTTACTGTGTCCAGAGCCCTATCT 1000  
|||||  
QY 961 aaaaagttggcatgaaagctgactccaggtcagcatgaaacgatggcctctctttttat 1020  
|||||  
DB 1001 AAAAGAAGTTGGCATGAAAGCTGCACCTCCAGGTCAGCATGAACGATGGCTCTCTTTAT 1060  
|||||  
QY 1021 ctccagttctgcatcatcaccacacacacacacacacacacacacacacacacacacacac 1080  
|||||  
DB 1061 CTCCAGTTCTGTGCATCATCACCACACACACACACTGTTCTGACGGTTCATCTTGCCCATCG 1120  
|||||  
QY 1081 cctgctgactgcttctctctctctctctctctctctctctctctctctctctctctctct 1140  
|||||  
DB 1121 CTTGCTGATCCTGTTCTCTCTCTAGCCCTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1180  
|||||  
QY 1141 ctgactgtgattatacaaggagggtccctccacccctgccaggagagagtgagaaaaa 1197  
|||||  
DB 1181 CTGCACTGTGATTATCAAGAGGTCCCTCCACCCCTGCCGAGGAGTGTGAGAGA 1237  
|||||  
  
RESULT 3  
BC012074 2112 bp mRNA linear PRI 06-AUG-2001  
LOCUS Homo sapiens, similar to tumor endothelial marker 8, clone  
DEFINITION MGC:19967 IMAGE:4563020, mRNA, complete cds.  
ACCESSION BC012074  
VERSION BC012074.1 GI:15082332  
KEYWORDS MGC.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2112)  
AUTHORS Strausberg,R.  
TITLE Direct Submission

Qy	1	agagccgcgaggaagccgcgcgagatgcgcgtccctgagggctgctggcgagttccgcg	60
Db	10	AGACCCGCGAGGAGAGGCCCGCGGATGGCGCTCCCTGAGGCTCGTGGCGAGTTCCGCG	69
Qy	61	agcgtgggaagagcgagaccctgtctcccgcggctcgcggccaatgcacgcggaagcg	120
Db	70	AGCGTGGGAAGAGCGAGGACCTGTCTCCCGCGGCTCGGSGCCATGCCACGCGGAGCG	129
Qy	121	gaagacccctcgacatcgcttccagtggtctctttggccactctgctgcctatcgcgc	180
Db	130	GAGAGCCCTCGGATCGGGTTCCAGTGGCTCTCTTGGGCCACTCTGGTGCATCTGC	189
Qy	181	cgaggcaagggggagcagggaggagatgggggtccagctctctacggcggaattgaacctgta	240
Db	190	CGGCAAGAGGGGACGACGAGGAGGATGGGGGTCCAGCCTGTACGGCGGATTTGACCTGTA	249
Qy	241	cttcattttgacaaatcaggaagtgctgcaccacatgggaatgaaatcattactttgt	300

Wakamatsu,

Crone, R. A. 1972. *Phylogenetic relationships of the Hominidae*.  
Homo sapiens  
Eukaryota; Metazoa  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (sites)  
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,  
Nishikawa, T., Nagai, K., Sugano, S., Takahashi-Fujii, A., Hara, H.,  
Tanase, T., Nomura, Y., Togiya, S., Komatsu, K., Takeuchi, K.,  
Arita, M., Nabekura, T., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J.,  
Wakamatsu, A., Nakamura, Y., Nagahari, K., Masuho, Y. and Osada, A.



TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished (2000)  
REFERENCE 2 (bases 1 to 1436)  
AUTHORS Isogai,T. and Otsuki,T.  
TITLE Direct Submission  
JOURNAL Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genomics@ri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing: Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing and clone selection:  
Helix Research Institute (supported by Japan Key Technology Center  
etc.) and Department of virology, Institute of Medical Science,  
University of Tokyo.

FEATURES  
source Location/Qualifiers  
1. 1436  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RP2005000"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2RP2"  
/note="cloning vector: pME18SFL3; mRNA from NT2 neuronal  
precursor cells after 2-weeks retinoic acid (RA)  
induction."  
380..1036  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAA91707.1"  
/db\_xref="GI:7022738"  
/translation="MSFVSTRTGTTLMKLTEDREQIROGLELQKVLPGDVTYMHSG  
FERASEQIYVENRGQYRTASVIALTDGELHEDLFFYSEREANRSLDGLAVYCVGK  
DFNETQLRADSKDHPVYNDGFQALQGIHSLKKSIEILAAEFTICAGESFYQ  
VVRNGEPRHARNDRVLCSPKINDSVLSKSLQSPWVSSTSGFGKEGNSHPCLPARPH  
T"

CDS  
380..1036  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAA91707.1"  
/db\_xref="GI:7022738"

BASE COUNT 375 a 349 c 401 g 311 t  
ORIGIN

Query Match 62.98; Score 890; DB 9; Length 1436;  
Best Local Similarity 99.3%; Pred. No. 1.2e-235;  
Matches 904; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Qy 1 aggcacgcgaggaagccgcgagatgcgctccctgaggtcgtagcttcgcgagttcgcg 60  
Db 41 AGGACCCCGAGAGAGGCCCGCGGATGGCGGTCCTCGAGGGCTGCGGAGTTGCGCG 100  
Qy 61 agcgtgggaagagcggaccctgtctccccgggctcgccgagccatggccacggcgagcg 120  
Db 101 ACGGTGGGAAGGAGCGGACCTGTCTCCCCGGGCTCGCGCCATGGCCACGGCGAGCG 160  
Qy 121 gagagccctcgccatcggtctccagtggtctctttggccacctctggtgctcatctgccc 180  
Db 161 GAGAGCCCTCGCATCGGCTTCCAGTGGCTCTC-ACGGCCACCTCTGTGCTCATCTGCGC 219  
Qy 181 cggcaggggagcagcagggagatgggggtccagcctgctacgcgagattgacctata 240  
Db 220 CGGCAAGGGGAGCGACGAGAGATGGGGGTCCACCTGCTACGGCGATTGACCTGTA 279  
Qy 241 ctctattttgacaaatcaggaagtgtgctgcaccactggaatgaaatctattattgt 300  
Db 280 CTTCAATTTTGACAAATCAGGAAGTGTGCTGCACCACCTGGAATGAAATCTATTACTTGT 339  
Qy 301 ggaacagtgtgctcacaattcaccagccacagttgagatgtccttatgtttctc 360  
Db 340 GGAACAGTTGGCTCACAATTCATCAGCCACAGTGTGAGATGTCTTATTGTTTCTC 399  
Qy 361 caccgaggaacaaccttaataagaaactacagagacagacacaaatccgtcaaggcct 420  
Db 400 CACCCGAGGACAACTTAATGAAGTACAGAGAGAGAGAAATCCCGTCAAGGCCT 459  
Qy 421 agaagaactccagaagaagtctgccaggaggagacacttacatgcatgaaggattgaaag 480

Db 460 AGAAGAACTCCAGAAAGTTCTGCCAGGAGGAGACACTTACATGCATGAAGATTGTAAG 519  
Qy 481 ggcagtgagcagatttattatgaacacagacagaggttacaggacagccagcgtcatcat 540  
Db 520 GGCCAGTGAGCAGATTTATTATGAACACAGACAAAGGTACAGGACGATGCGTCATCAT 579  
Qy 541 tgccttgaactgaggaacacccatgaagatctctttctattcagagagggagcctaa 600  
Db 580 TGCCTTGACTGATGGAGAACTCCATGAAGATCTCTTTTCTATTTCAGAGAGGGAGGCTAA 639  
Qy 601 taggtctcagatcttgctgcaattgtttactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 660  
Db 640 TAGGTCTCGACATCTGTGCAATTTGTTTACTGTGTGCTGTGTAAGATTTCAATGACAC 699  
Qy 661 acagtgcccgagattgcggacagtaagatcatgtgtttcccgtaagatgacgggttca 720  
Db 700 ACAGTGGCCCGGATTCGGACAGTAAGGATCATGTGTTTCCCGTGAATGACGGCTTTCA 759  
Qy 721 ggcctgcaagcgcacatccactcaattttgaagaagtcctgcatcgcaaatctctagcgc 780  
Db 760 GGCTCTGCAAGGCATCATCCACTCAATTTTGAAGAAGTCTCTGCATCGAAATTTCTAGCAGC 819  
Qy 781 tgaaccatccaccatattgtcaggagatcatttcaagtgtgctgtagagaaacggctt 840  
Db 820 TGAACCATCCACCATATGTGCGAGGAGAGTCAATTTCAAGTTGTGCTGAGAGGAACGGCTT 879  
Qy 841 ccgacatcccgcaacgctggacaggggtcctctgcaagctcgaagatcaaatgacgtcgctac 900  
Db 880 CCGACATGCCCGCAACGCTGGACAGGCTCCTCTGCAGTTCATGATCAATGACTCGGTCCAC 939  
Qy 901 actcaatgag 910  
Db 940 ACTCAGTAAG 949

RESULT 5  
AF378762 5220 bp mRNA linear ROD 07-OCT-2001  
LOCUS AF378762 Mus musculus tumor endothelial marker 8 precursor (Tem8) mRNA,  
DEFINITION complete cds.  
ACCESSION AF378762  
VERSION AF378762.1 GI:15987504  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 5220)  
AUTHORS Carson-Walter,E.B., Vogelstein,B., Kinzler,K.W. and St. Croix,B.  
TITLE Cell surface tumor endothelial markers are conserved in mice and humans  
JOURNAL Cancer Res. 61 (18), 6649-6655 (2001)  
MEDLINE 21443268  
PUBMED 11559528  
REFERENCE 2 (bases 1 to 5220)  
AUTHORS Carson-Walter,E.B., Vogelstein,B., Kinzler,K.W. and St. Croix,B.  
TITLE Direct Submission  
JOURNAL Submitted (10-MAY-2001) Oncology, Johns Hopkins University, 1650  
Orleans Street, Baltimore, MD 21231, USA  
FEATURES  
source Location/Qualifiers  
1. 5220  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
1. 5220  
/gene="Tem8"  
274..1962  
/gene="Tem8"  
/note="mTEM8"  
/codon\_start=1  
/product="tumor endothelial marker 8 precursor"  
/protein\_id="AAL11995.1"

```
/db_xref="GI:15987505"
/translation="MDRAGLGAAGLGLCAALVLCVAGHGRRGDPACVGGFDLY
FLDKSGSVLHWNHIIYFEQALAHRFISPOGRSFIYFSTRTGLMLKLTREDRIQ
GLEELQKVLGGDTYMHGFERASEQIYVENSQGYRTASVIALJDGELHDFIYSE
REANRSRLGATVYGVKDFNETQIARIADSKDHVFPVNDQFQALQIHLKSKC
IETLAEPSTICAGSFQVVRNGFRHARNRDLVLCSPKINDSVTLNEKPFPAVEDTY
LLCPAPILKEVGMKAALOVSMNDGSLFISSVIITTHCSDGSLAIALLVFLLLAL
ALWFWPLCCVIIKEVPPPPVSESEEDDGLPKKWPVTVDASYVGGVGGIKRM
EVRWGKSTEGAKLEKARNKARVMPDEYEFPEPRNLNNMRPSPSRKWSPIKG
KLDALVLLRGDYSVNMROPQDGTGRICINFRVKNRSPAKYPLNNIYTHPSPPAP
IYTPPPAPHPCHPPAPAPPTPIPPSPPLPPPPQAPPPNRPAPPPSPRPPPSV"
274..354
sig_peptide
/gene="Tem8"
355..1959
mat_peptide
/product="tumor endothelial marker 8"
1228..1293
misc_feature
/gene="Tem8"
/notes="Region: predicted transmembrane domain"
BASE COUNT 1406 a 1329 c 1235 g 1250 t
ORIGIN
Query Match 60.4%; Score 853.4; DB 10; Length 5220;
Best Local Similarity 87.3%; Pred. No. 1e-215;
Matches 935; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
Qy 127 cctcggaatggttcacagtggtctctcttggccactctgggtgctcattcgcgcggaca 186
Db 291 CCTGGTGGCGGCTCGGGGACTCTCGTGGCTGCACTCGTGTGTCGTGGCGGGGACA 350
Qy 187 aggggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 246
Db 351 CGGGGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 410
Qy 247 ttggacaaatcagaagtgctgcaccactggaatggaatctattacttttgggaaca 306
Db 411 CCTGCACAAATCAGAAAGTGTCTGCACCCTGGAATGAAATCTACTACTTCTGCGGAGCA 470
Qy 307 gttggtcacaaatcaccacacacacacacacacacacacacacacacacacacacacac 366
Db 471 GTTGGCTCATAGATTCAATCAGCCACACAGTAAAGATGTCCTTCATGTGCTTCTCTACWCG 530
Qy 367 aggaacaacctaatgaactgaacagacagacagacacacacacacacacacacacacacac 426
Db 531 AGGGACACTTAAATGAACATTAAGTGAAGGACAGGACAGGACAGGACAGGACAGGACAGG 590
Qy 427 actccagaaagtctgcagaggagacacattacatgcatgaagatttgaaggccag 486
Db 591 GCTCCAGAAAGTTCTCCAGGAGGAGACACTTACATGACCAAGGATTCGAGAGGGCCAG 650
Qy 487 tgagcagatttattataaagacagacagcagcagcagcagcagcagcagcagcagcagcagc 546
Db 651 TGAGCAGATTCTACTATGACACAGTCAAGATACAGGAGGAGGAGGAGGAGGAGGAGGAGG 710
Qy 547 gactgatgagacacccatgaagatctcttttctattcagagagggaggttaatagtc 606
Db 711 GACGGATGGGAGCTGCAGGAGGAGCTTCTTCTACTCAGAGAGGAGGAGGAGGAGGAGGAG 770
Qy 607 tcgagatttggatgaatttatttactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 666
Db 771 CCGAGACCTTGTGGGATGTTTACTGGCTGGCGTGAAGGATTCAATGAACATCAGTT 830
Qy 667 gcccggattgcgagcagtaagatcattgtttcccgtaagtcagcggcttcaggctct 726
Db 831 GCCTCGGATTGACAGACAGTAAAGACACAGTGTTCCTGTGACGACGAGGCTTCAGGCTCT 890
Qy 727 gcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 786
Db 891 CCAAGGCATTATCCATCAATTTAAAGAAATCTGCATCGAAATCTGGCGGCTGAACC 950
Qy 787 atccaccatgtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 846
```

```
Db 951 ATCCACCATCTCGCGGGAGAGTCCTTTCAAGTGTGTAAGAGGAAATGGCTTCGGACA 1010
Qy 847 tcccgcacagtgggacagaggtctctctgcagctcaagatcaatgactcggtcacactcaa 906
Db 1011 TGCCCGCATGTGGACAGGGTCTCTGCAGCTCAAAATCAATGACTCAGTCACGCTCAA 1070
Qy 907 tgagaagccctttctgtgtgaagacacttatttactgtgtccagcgcctattttaaaga 966
Db 1071 TGAGAAGCCCTTTGCTGTGTGAAGACACTTATTGCTGTGCCAGCACCAATCTTTGAAGA 1130
Qy 967 agttgcataaagctgcactcagctcagcatgaacgatggcctctcttttatctccag 1026
Db 1131 AGTTGGCATGAAGGCTGCACCTGCAGGTGACATGAACGACGGCTGTCTTCATCTCCAG 1190
Qy 1027 ttctgcatcatcaccaccacacacactgttctgacggttccatcctggccatgcgctgt 1086
Db 1191 TTCTGTCAATCATCACCACCACACACTGTTTCAGACGGCTCCATCTCGCGATTCTCTGCT 1250
Qy 1087 gactctgttctgctcactgagcctgctctctctgtgtgtgtgtgtgtgtgtgtgtgtgt 1146
Db 1251 GGTCTCTTCTCTGCTGCTGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1310
Qy 1147 tgtgattatcaagaggtccctccacccctgcgagagagtgaggaaga 1197
Db 1311 AGTATCATCAAGAGGTCCCTCCACCCCTGTTGAGGAGAGTGAGGAAGA 1361
RESULT 6
AK025429 Homo sapiens cDNA: FLJ21776 fis, clone HEP00171. linear
LOCUS AK025429
DEFINITION AK025429
ACCESSION AK025429
VERSION AK025429.1 GI:10437938
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens hepatoma cell_line:HepG2 cDNA to mRNA, clone_lib:HEP
clone:HEP00171.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished (2000)
JOURNAL 2 (bases 1 to 4007)
REFERENCE Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
AUTHORS Direct Submission
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES Location/Qualifiers
source 1. 4007
/db_xref="taxon:9606"
/clone="HEP00171"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone_lib="HEP"
/notes="cloning vector pME18SFL3"
323..1147
/notes="unnamed protein product"
/codon_start=1
CDS
```

```

/protein_id="BAB15128.1"
/db_xref="GI:10437939"
/translation="MKAALQYSNMDGLSFISSSVLIITTHCSOGSIIAIIALLILFLLL
ALALLWFJLCTTVIKVEYPPPAESEDDEDDGLPKRWPTVDASYIGRGVGGIK
RMEVWKEGKSTBEGAKLEKAKARVMKPEOYEFPEPRNLNNMRPSPRWKWSPI
KGLDLMLVLLRGYDRVSMRQPGDTGRINFTRKNOPAKYPLNNAYHTSSPPP
APIYTPPPPHCPPPPPSPAPTPIPSPPSTLPPPPQAPPPNRPAPPPSRPPRPVS"
BASE COUNT      1144 a 1003 c 832 g 1028 t
ORIGIN

```

```

Query Match      38.4%; Score 542.8; DB 9; Length 4007;
Best Local Similarity 99.6%; Pred. No. 4.2e-133;
Matches 544; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 652 caatgagacacagctgcccagattgagacagtaagatcatgtgttccgtgaatga 711
Db 1 CAATGAGACACAGTGGCCGGATGGCGACAGTAAGATCATGTGTTCCTCCGTAATGA 60
Qy 712 cggctttcaggctctgcaaggcatcatccactcaatttgaagagtcctgcacgaat 771
Db 61 CGGCTTCAGGCTCTGCAAGGCATCATCTCAATTTTGAGAGTCTCTGCATCGAAT 120
Qy 772 tctagcagctgaaccatccaccatgtgtaggagagtcatttcaagtgtgctgagagg 831
Db 121 TCTAGCAGCTGAACCATCCACATATGTGCAGGAGTCAATTTCAAGTTGTCGTGAGAGG 180
Qy 832 aaacggcttcgacatgcccgaacgtagacagggctctgcagcttcaagataatga 891
Db 181 AAACGGCTTCGACATGCCCGCAACGTCGACAGAGTCTCTGCAAGTTCGAAGATCAATGA 240
Qy 892 ctggtcacactcaatgagagccctttctgtggaagacacttatttactgtgtccagc 951
Db 241 CTCGGTCACTCAATGAGAGACCTTTTCTGTGGAAATATCTTATTTACTGTGTGCCAGC 300
Qy 952 gcctattctaaagaagtgtgcatgaaagctgactccagctcagctcagctgaacgatggcct 1011
Db 301 GCCTATCTTAAAGAAGTGTGGCATGAAAGCTGCACCTCCAGGTCAGCATGAACGATGGCCT 360
Qy 1012 ctctttatctccagttctgcatcatcaacacacacacactgttctgagcgttccatcct 1071
Db 361 CTCCTTTATCTCCAGTCTGTGCATCATCACCACACACACTGTTCTGACGGTTCCATCCT 420
Qy 1072 ggcacgctcctgctgacccgttctctgctcctagccctgctcctcctctggtgttctg 1131
Db 421 GGCCATCGCCCTGCTGATCCTGCTCTCTAGCCCTGGCTCTCCCTCTGGTGGTCTG 480
Qy 1132 gcccctctgctgactgtgtatatacgaaggaggtccctccacccctgcccaggagagtgga 1191
Db 481 GCCCTCTGCTGCACTGTGATTATCAAGGAGTCCCTCCACCCCTGCCGAGAGAGTGA 540
Qy 1192 ggaataa 1197
Db 541 GGAAGA 546

```

```

RESULT 7
LOCUS AK057316 2126 bp mRNA linear PRI 31-OCT-2001
DEFINITION Homo sapiens cDNA FLJ32754 fis, clone TEST12001671.
ACCESSION AK057316
VERSION AK057316.1 GI:16552959
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens testis cDNA to mRNA, clone_lib:TEST12
clone:TEST12001671.

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S.,
Ono,Y., Hotuta,T., Hiroaka,S., Murakawa,K., Takiguchi,S.,
Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M.,
Yamashita,H., Chiba,Y., Sugiyama,T., Irie,R., Otsuki,T., Sato,H.,

```

```

Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y.,
Saito,K., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y.,
Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Takahashi-Fujii,A.,
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2126)
Isogai,T., Otsuki,T. and Sugiyama,T.
Direct Submission
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan: cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: RAB and
HRI.

```

```

FEATURES             Location/Qualifiers
     source            1..2126
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="TEST12001671"
                     /tissue_type="testis"
                     /clone_lib="TEST12"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      536 a 614 c 517 g 459 t
ORIGIN

```

```

Query Match      18.1%; Score 255.8; DB 9; Length 2126;
Best Local Similarity 56.6%; Pred. No. 7.7e-57;
Matches 495; Conservative 0; Mismatches 377; Indels 3; Gaps 1;

Qy 183 ggcgaaggggcgagggaggggtccagcctctagcggggttgaacctgtact 242
Db 143 GGCACACTGGCCGAGGAGGAGGTGCACAGATGCGAGGCTCATTTGACCTCTACT 202
Qy 243 tcatttggacaaatcagggaagtgtgctgcaccactggaatgaatcttacttttgg 302
Db 203 TCATCTTGGCAAGTCTGGCAGCGTGAACAATAATGATGGATTGACCTTTATATATGGGTGG 262
Qy 303 aacagttggtcacaaattcatcagccacacagttgagaatgtccttttctttctcca 362
Db 263 AGGAACACAGTGGCGAGGTTCCAAAGCCCCAAATATTTCGGATGTGCTTCACCTACTCCA 322
Qy 363 cccgagggaacaaccttaataagactgacagaagacagagacacaaatccgtcaagcctag 422
Db 323 CAGACGGCCAGACTGCTCTTGCCACTCAGCTCAGACAGAAATAGATAAAAAACGGCTCTTG 382
Qy 423 aagaactccagaagaagtctctgcaggaggagacacttacatgcattgaaggattgaagg 482
Db 383 ACCAACTTCAGAAAAATTGTGCTGACGGTACACACTTTCAGGAGGAGGATTAGAAAGG 442
Qy 483 ccagtgaacagatttatatgaagaacagaaaggtacaggaacagccagcgcctcatgtg 542
Db 443 CAATTCAACAGATCCAAAATTTTCAAC---TCGGGAAACAAGGTTCCAGCATGATTATTG 499
Qy 543 ctttgactgaggagaactcctcatgaagatctcttcttcttcttcttcttcttcttctt 602
Db 500 CTATGACTGATGGAGAACTGGTGGCACATGCATTTTCAGGACACTCTCAGAGAAAGTCAAA 559
Qy 603 ggtctcagagatcttgggtcaattgtttactgtgttggtgtgaaagatttcaatgagacac 662
Db 560 AGGCTCGGAAACTGGGGCCAAACGTTTACACCTCGGTGTGGCTGATTAATATCTGGAGC 619
Qy 663 agctggcccgagattcgagacagtaaggatcatgtgtttctccctggaatgacgctttcagg 722
Db 620 AGATAACAGCAATTTCAGACACGCCCTGGCCAGCTGCTTTTGCAGTGAGAAATGCGCTTCAAGG 679

```





Db 321 GAAATTTATAATTCGTACAGCAACTTGGGAGAGAGATTTGTGAGCCCTGAATTCAGATTGA 380

Qy 344 tctttattgttttcccccagagaaacaccttaattgaactgacagaagacagagaa 403

Db 381 TCATTTCATGTTCTTCTTCAAGCAACTTATTTTGGCCATTAACCTGGAGACAGAGGC 440

Qy 404 caaatccgtcaagccttagaagaactccagaaagttctgcagagagagacacttaccatg 463

Db 441 AAAATCAGTAAGAGCTGGAGAGATTTAAAACGTTAGTCCAGTAGGAGACATATATC 500

Qy 464 catgaagattgaagaggccagtgcagacatattattatgaacacagacaaggggtacagg 523

Db 501 CATGAAGGACTAAAGCTAGCGAATGACAAAT-----TCAGAAAGCAGAGGCTTGAAA 554

Qy 524 acagccagcgtcatcattgttgaactgtaggagaactccatccatgaagatctcttttttat 593

Db 555 ACCTCCAGTATCATATTTGCTCTGACAGATGGCAAGTTGGACGCTCTGTGCCATCATAT 614

Qy 584 tcagagaggggaggaataaggtctcgagatcttggtcaattgttactgttgggtg 643

Db 615 GCAGAGAAAGAGGCAAGATATCCAGTCTAGTGGGCTAGTGTATTTGTGTGGTCTC 674

Qy 644 aaagatttcaatgagacacagctgcccggattgctgcagacagtaggaagatcatgtgttccc 703

Db 675 CTTCATTTGACAAAGCACACCTTGAAAGAAATTCGATCCCAAGGAGCAAGTTTCCCT 734

Qy 704 gtgaatgacggttcaggctctcgaaggaatcatcacc 744

Db 735 GTCAGAGGTGATTCAGGCTCTTAAAGGAATAATTAATC 775

RESULT 11

AC025010 81017 bp DNA linear HTG 13-JUL-2000

LOCUS Homo sapiens chromosome 4 clone RP11-85D17 map 4, LOW-PASS SEQUENCE

DEFINITION

ACCESSION AC025010

VERSION AC025010.1 GI:7145054

KEYWORDS HTG; HTGS\_PHASE0.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE 1 (bases 1 to 81017)

JOURNAL Homo sapiens chromosome 4, clone RP11-85D17

REFERENCE 2 (bases 1 to 81017)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castile, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Glende, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Hearford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Melrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE Direct Submission

JOURNAL Submitted (03-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: L7756  
 Center clone name: 85\_D\_17

\*\*\* NOTE: This record contains 94 individual  
 \*\*\* sequencing reads that have not been assembled into  
 \*\*\* contigs. Runs of N are used to separate the reads  
 \*\*\* and the order in which they appear is completely  
 \*\*\* arbitrary. Low-pass sequence sampling is useful for  
 \*\*\* identifying clones that may be gene-rich and allows  
 \*\*\* overlap relationships among clones to be deduced.  
 \*\*\* However, it should not be assumed that this clone  
 \*\*\* will be sequenced to completion. In the event that  
 \*\*\* the record is updated, the accession number will  
 \*\*\* be preserved.

1 709: contig of 709 bp in length  
 \* 710 809: gap of 100 bp  
 \* 810 1567: contig of 758 bp in length  
 \* 1568 1667: gap of 100 bp  
 \* 1668 2433: contig of 766 bp in length  
 \* 2434 2533: gap of 100 bp  
 \* 2534 3310: contig of 777 bp in length  
 \* 3311 3410: gap of 100 bp  
 \* 3411 4176: contig of 766 bp in length  
 \* 4177 4276: gap of 100 bp  
 \* 4277 5036: contig of 760 bp in length  
 \* 5037 5136: gap of 100 bp  
 \* 5137 5899: contig of 763 bp in length  
 \* 5900 5999: gap of 100 bp  
 \* 6000 6764: contig of 765 bp in length  
 \* 6765 6864: gap of 100 bp  
 \* 6865 7612: contig of 748 bp in length  
 \* 7613 7712: gap of 100 bp  
 \* 7713 8467: contig of 755 bp in length  
 \* 8468 8567: gap of 100 bp  
 \* 8568 9320: contig of 753 bp in length  
 \* 9321 9420: gap of 100 bp  
 \* 9421 10187: contig of 767 bp in length  
 \* 10188 10287: gap of 100 bp  
 \* 10288 11042: contig of 755 bp in length  
 \* 11043 11142: gap of 100 bp  
 \* 11143 11910: contig of 768 bp in length  
 \* 11911 12010: gap of 100 bp  
 \* 12011 12757: contig of 747 bp in length  
 \* 12758 12857: gap of 100 bp  
 \* 12858 13619: contig of 762 bp in length  
 \* 13620 13719: gap of 100 bp  
 \* 13720 14493: contig of 774 bp in length  
 \* 14494 14593: gap of 100 bp  
 \* 14594 15348: contig of 755 bp in length  
 \* 15349 15448: gap of 100 bp  
 \* 15449 16208: contig of 758 bp in length  
 \* 16207 16307: gap of 100 bp  
 \* 16307 17054: contig of 748 bp in length  
 \* 17055 17154: gap of 100 bp  
 \* 17155 17900: contig of 746 bp in length  
 \* 17901 18000: gap of 100 bp  
 \* 18001 18767: contig of 767 bp in length  
 \* 18768 18867: gap of 100 bp  
 \* 18868 19642: contig of 775 bp in length  
 \* 19643 19742: gap of 100 bp  
 \* 19743 20493: contig of 751 bp in length  
 \* 20494 20593: gap of 100 bp  
 \* 20594 21337: contig of 744 bp in length  
 \* 21338 21437: gap of 100 bp  
 \* 21438 22202: contig of 765 bp in length



Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (03-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L7756  
Center clone name: 85\_D\_17

NOTE: This record contains 94 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved. 709: contig of 709 bp in length

1  
\* 710 809: gap of 100 bp  
\* 810 1567: contig of 758 bp in length  
\* 1568 1667: gap of 100 bp  
\* 1668 2433: contig of 766 bp in length  
\* 2434 2533: gap of 100 bp  
\* 2534 3310: contig of 777 bp in length  
\* 3311 3410: gap of 100 bp  
\* 3411 4176: contig of 766 bp in length  
\* 4177 4276: gap of 100 bp  
\* 4277 5036: contig of 760 bp in length  
\* 5037 5136: gap of 100 bp  
\* 5137 5899: contig of 763 bp in length  
\* 5900 5999: gap of 100 bp  
\* 6000 6764: contig of 765 bp in length  
\* 6765 6864: gap of 100 bp  
\* 6865 7612: contig of 748 bp in length  
\* 7613 7712: gap of 100 bp  
\* 7713 8467: contig of 755 bp in length  
\* 8468 8567: gap of 100 bp  
\* 8568 9320: contig of 753 bp in length  
\* 9321 9420: gap of 100 bp  
\* 9421 10187: contig of 767 bp in length  
\* 10188 10287: gap of 100 bp  
\* 10288 11042: contig of 755 bp in length  
\* 11043 11142: gap of 100 bp  
\* 11143 11910: contig of 768 bp in length  
\* 11911 12010: gap of 100 bp  
\* 12011 12757: contig of 747 bp in length  
\* 12758 12857: gap of 100 bp  
\* 12858 13619: contig of 762 bp in length  
\* 13620 13719: gap of 100 bp  
\* 13720 14493: contig of 774 bp in length  
\* 14494 14593: gap of 100 bp  
\* 14594 15348: contig of 755 bp in length  
\* 15349 15448: gap of 100 bp  
\* 15449 16206: contig of 758 bp in length  
\* 16207 16306: gap of 100 bp  
\* 16307 17054: contig of 748 bp in length  
\* 17055 17154: gap of 100 bp  
\* 17155 17900: contig of 746 bp in length  
\* 17901 18000: gap of 100 bp  
\* 18001 18767: contig of 767 bp in length  
\* 18768 18867: gap of 100 bp  
\* 18868 19642: contig of 775 bp in length  
\* 19643 19742: gap of 100 bp

19743 20493: contig of 751 bp in length  
20494 20593: gap of 100 bp  
20594 21337: contig of 744 bp in length  
21338 21437: gap of 100 bp  
21438 22202: contig of 765 bp in length  
22203 22302: gap of 100 bp  
22303 23072: contig of 770 bp in length  
23073 23172: gap of 100 bp  
23173 23941: contig of 769 bp in length  
23942 24041: gap of 100 bp  
24042 24799: contig of 758 bp in length  
24800 24899: gap of 100 bp  
24900 25667: contig of 768 bp in length  
25668 25767: gap of 100 bp  
25768 26515: contig of 748 bp in length  
26516 26615: gap of 100 bp  
26616 27375: contig of 760 bp in length  
27376 27475: gap of 100 bp  
27476 28238: contig of 763 bp in length  
28239 28338: gap of 100 bp  
28339 29104: contig of 766 bp in length  
29105 29204: gap of 100 bp  
29205 29969: contig of 765 bp in length  
29970 30069: gap of 100 bp  
30070 30837: contig of 768 bp in length  
30838 30937: gap of 100 bp  
30938 31710: contig of 773 bp in length  
31711 31810: gap of 100 bp  
31811 32576: contig of 766 bp in length  
32577 32676: gap of 100 bp  
32677 33461: contig of 785 bp in length  
33462 33561: gap of 100 bp  
33562 34337: contig of 776 bp in length  
34338 34437: gap of 100 bp  
34438 35216: contig of 779 bp in length  
35217 35316: gap of 100 bp  
35317 36082: contig of 766 bp in length  
36083 36182: gap of 100 bp  
36183 36951: contig of 769 bp in length  
36952 37051: gap of 100 bp  
37052 37815: contig of 764 bp in length  
37816 37915: gap of 100 bp  
37916 38676: contig of 761 bp in length  
38677 38776: gap of 100 bp  
38777 39563: contig of 787 bp in length  
39564 39663: gap of 100 bp  
39664 40445: contig of 782 bp in length  
40446 40545: gap of 100 bp  
40546 41316: contig of 771 bp in length  
41317 41416: gap of 100 bp  
41417 42181: contig of 765 bp in length  
42182 42281: gap of 100 bp  
42282 43030: contig of 749 bp in length  
43031 43130: gap of 100 bp  
43131 43913: contig of 783 bp in length  
43914 44013: gap of 100 bp  
44014 44770: contig of 757 bp in length  
44771 44870: gap of 100 bp  
44871 45631: contig of 761 bp in length  
45632 45731: gap of 100 bp  
45732 46487: contig of 756 bp in length  
46488 46587: gap of 100 bp  
46588 47340: contig of 753 bp in length  
47341 47440: gap of 100 bp  
47441 48171: contig of 731 bp in length  
48172 48271: gap of 100 bp  
48272 49028: contig of 757 bp in length  
49029 49128: gap of 100 bp  
49129 49883: contig of 755 bp in length  
49884 49983: gap of 100 bp  
49984 50748: contig of 765 bp in length  
50749 50848: gap of 100 bp  
50849 51619: contig of 771 bp in length



JOURNAL	unpublished (2000)
REFERENCE	2 (bases 1 to 1189)
AUTHORS	Isogai,T. and Otsuki,T.
TITLE	Direct Submission
JOURNAL	Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:01-438-52-3951, Fax:01-438-52-3952)
COMMENT	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library

```
1. 2557
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3484366"
/tissue_type="Mammary tu
old, gross tissue."
```

COMMENT	<p>NEDO human CDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library</p> <p>/cd_xref="taxon:10090 /clone=IMAGE:3484366" /tissue_type="Mammary old, gross tissue."</p> <p>WAP-tGF alpha model. 7 months</p>
---------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------



Search completed: August 9, 2002, 00:23:20  
Job time: 12457 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:35:58 ; Search time 98.32 Seconds  
(without alignments)  
388.682 Million cell updates/sec

Title: US-09-970-076-10  
Perfect score: 1131  
Sequence: 1 MSFVSTRTGTTLMKLTEDR.....STSGFKEGNSHPCLPARPHT 218

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 528882 seqs, 175299045 residues

Total number of hits satisfying chosen parameters: 528882

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_New.\*

- 1: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1131	100.0	218	5	US-09-629-469A-11706
2	966	85.4	328	6	US-10-038-307-26
3	966	85.4	333	6	US-10-038-307-2
4	966	85.4	342	6	US-10-038-307-22
5	966	85.4	345	6	US-10-038-307-24
6	966	85.4	551	6	US-10-038-307-18
7	966	85.4	564	1	PCT-US02-08253-187
8	966	85.4	564	1	PCT-US02-08253-232
9	966	85.4	564	6	US-10-038-307-20
10	966	85.4	564	7	US-60-389-987-1823
11	960	84.9	562	1	PCT-US02-08253-194
12	960	84.9	562	1	PCT-US02-08253-301
13	945	83.6	534	6	US-10-038-307-12
14	945	83.6	543	6	US-10-038-307-10
15	945	83.6	543	6	US-10-038-307-14
16	945	83.6	543	6	US-10-038-307-16
17	704	62.2	538	6	US-10-047-542-99
18	506	44.7	488	6	US-10-104-047-2639
19	506	44.7	488	6	US-10-038-307-6
20	506	44.7	488	7	US-60-373-595-10
21	505	44.7	587	6	US-10-125-540-312
22	495	43.8	487	6	US-10-038-307-8
23	415	36.7	381	6	US-10-038-307-4
24	111	9.8	1152	5	US-09-592-617A-43
25	105.5	9.3	3063	1	PCT-US02-08253-257
26	105.5	9.3	3063	1	PCT-US02-19669-61

27	105.5	9.3	3063	1	PCT-US02-19669-63
28	105.5	9.3	3063	6	US-10-177-293-61
29	105.5	9.3	3063	6	US-10-177-293-63
30	99.5	8.8	460	6	US-10-153-881-33525
31	99.5	8.8	646	6	US-10-155-881-22380
32	98	8.7	191	5	US-09-805-354-1
33	98	8.7	1163	1	PCT-US02-10824-116
34	98	8.7	1163	7	US-60-369-452-33
35	96.5	8.5	832	7	US-60-360-039-2170
36	95	8.4	764	1	PCT-US02-19669-27
37	95	8.4	764	6	US-10-177-293-27
38	92.5	8.2	4307	7	US-60-360-039-5698
39	92.5	8.2	4307	7	US-60-360-039-5699
40	92.5	8.2	4307	7	US-60-360-039-5700
41	92	8.1	915	5	US-09-904-920A-34
42	92	8.1	915	6	US-10-121-049-294
43	92	8.1	915	6	US-10-121-050-294
44	92	8.1	915	6	US-10-121-053-294
45	92	8.1	915	6	US-10-121-043-294

## ALIGNMENTS

RESULT 1  
US-09-629-469A-11706  
; Sequence 11706, Application US/09629469A  
; GENERAL INFORMATION:  
; APPLICANT: OTA, TOSHIO  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: NISHIKAWA, TETSUO  
; APPLICANT: HAYASHI, KOJI  
; APPLICANT: SAITO, KAORU  
; APPLICANT: YAMAMOTO, JUNICHI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: OTSUKI, TETSUJI  
; TITLE OF INVENTION: PRIMERS FOR SYNTHESIZING FULL-LENGTH CDNA AND THEIR USE  
; FILE REFERENCE: 084335/0123  
; CURRENT APPLICATION NUMBER: US/09/629,469A  
; CURRENT FILING DATE: 2000-07-28  
; PRIOR APPLICATION NUMBER: JP 1999-248036  
; PRIOR FILING DATE: 1999-07-29  
; PRIOR APPLICATION NUMBER: JP 1999-300253  
; PRIOR FILING DATE: 1999-08-27  
; PRIOR APPLICATION NUMBER: JP 2000-118776  
; PRIOR FILING DATE: 2000-01-11  
; PRIOR APPLICATION NUMBER: JP 2000-183767  
; PRIOR FILING DATE: 2000-05-02  
; PRIOR APPLICATION NUMBER: JP 2000-241899  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: 60/159,590  
; PRIOR FILING DATE: 1999-10-18  
; PRIOR APPLICATION NUMBER: 60/183,322  
; PRIOR FILING DATE: 2000-02-17  
; NUMBER OF SEQ ID NOS: 19025  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 11706  
; LENGTH: 218  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-629-469A-11706

Query Match 100.0%; Score 1131; DB 5; Length 218;  
Best Local Similarity 100.0%; Pred. No. 5.6e-106;  
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSFVSTRTGTTLMKLTEDREQIRQGLEELQKVLPGDGYMHEGFRASEQIYYENRQGY 60  
DB 1 MSFVSTRTGTTLMKLTEDREQIRQGLEELQKVLPGDGYMHEGFRASEQIYYENRQGY 60

```
; ORGANISM: Homo sapiens
US-10-038-307-2

Query Match      85.4%; Score 966; DB 6; Length 333;
Best Local Similarity 98.9%; Pred. No. 4.2e-89;
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFIVFSTRGTTIMKLTEDREQIRQGLEELQKVLPGGDTYMHGFFERASEQIYYENRQGY 60
Db 80 MSFIVFSTRGTTIMKLTEDREQIRQGLEELQKVLPGGDTYMHGFFERASEQIYYENRQGY 139
QY 61 RTASVIIALTGDGELHEDLFFYSEREANRSDLGAIYVCVKDFNETQIARIADSKDHVF 120
Db 140 RTASVIIALTGDGELHEDLFFYSEREANRSDLGAIYVCVKDFNETQIARIADSKDHVF 199
QY 121 PVNDGFOALQGIHSLKKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSE 180
Db 200 PVNDGFOALQGIHSLKKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSE 259
QY 181 KINDSVTLSK 190
Db 260 KINDSVTLNE 269

RESULT 4
US-10-038-307-22
; Sequence 22, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-26

Query Match      85.4%; Score 966; DB 6; Length 328;
Best Local Similarity 98.9%; Pred. No. 4.2e-89;
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFIVFSTRGTTIMKLTEDREQIRQGLEELQKVLPGGDTYMHGFFERASEQIYYENRQGY 60
Db 80 MSFIVFSTRGTTIMKLTEDREQIRQGLEELQKVLPGGDTYMHGFFERASEQIYYENRQGY 139
QY 61 RTASVIIALTGDGELHEDLFFYSEREANRSDLGAIYVCVKDFNETQIARIADSKDHVF 120
Db 140 RTASVIIALTGDGELHEDLFFYSEREANRSDLGAIYVCVKDFNETQIARIADSKDHVF 199
QY 121 PVNDGFOALQGIHSLKKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSE 180
Db 200 PVNDGFOALQGIHSLKKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSE 259
QY 181 KINDSVTLSK 190
Db 260 KINDSVTLNE 269

RESULT 3
US-10-038-307-2
; Sequence 2, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 333
; TYPE: PRT
```

```
; APPLICANT: Engin Ozkaynak
; APPLICANT: Judith J. Healey
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-24

Query Match      85.4%; Score 966; DB 6; Length 345;
Best Local Similarity 98.9%; Pred. No. 4.4e-89;
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFVIFSTRGTTLMKLTEDREIQROGLELQKVLPGGDTYMHGEGFERASEQIYYENRQGY 60
DB 80 MSFVIFSTRGTTLMKLTEDREIQROGLELQKVLPGGDTYMHGEGFERASEQIYYENRQGY 139
QY 61 RTASVITIALTDGELHEDLFFYSEREANRSRDGLGAIYVCVGVKDFNETQLARIADSKDHVF 120
DB 140 RTASVITIALTDGELHEDLFFYSEREANRSRDGLGAIYVCVGVKDFNETQLARIADSKDHVF 199
QY 121 PVNDGFOALQGIHHSILKSKCIEILAAEPSTICAGESFQVYVVRGNGFRHARNVDRVLCSF 180
DB 200 PVNDGFOALQGIHHSILKSKCIEILAAEPSTICAGESFQVYVVRGNGFRHARNVDRVLCSF 259
QY 181 KINDSVTLSK 190
DB 260 KINDSVTLINE 269

RESULT 6
US-10-038-307-18
; Sequence 18, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. Rottman
; APPLICANT: Theresa L. O'Keefe
; APPLICANT: Engin Ozkaynak
; APPLICANT: Judith J. Healey
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-18

Query Match      85.4%; Score 966; DB 6; Length 551;
Best Local Similarity 98.9%; Pred. No. 8.4e-89;
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFVIFSTRGTTLMKLTEDREIQROGLELQKVLPGGDTYMHGEGFERASEQIYYENRQGY 60
DB 80 MSFVIFSTRGTTLMKLTEDREIQROGLELQKVLPGGDTYMHGEGFERASEQIYYENRQGY 139
QY 61 RTASVITIALTDGELHEDLFFYSEREANRSRDGLGAIYVCVGVKDFNETQLARIADSKDHVF 120
DB 140 RTASVITIALTDGELHEDLFFYSEREANRSRDGLGAIYVCVGVKDFNETQLARIADSKDHVF 199
QY 121 PVNDGFOALQGIHHSILKSKCIEILAAEPSTICAGESFQVYVVRGNGFRHARNVDRVLCSF 180
DB 200 PVNDGFOALQGIHHSILKSKCIEILAAEPSTICAGESFQVYVVRGNGFRHARNVDRVLCSF 259
QY 181 KINDSVTLSK 190
DB 260 KINDSVTLINE 269
```

```
QY 181 KINDSVTLSK 190
DB 260 KINDSVTLINE 269

RESULT 7
PCT-US02-08253-187
; Sequence 187, Application PC/TUS0208253
; GENERAL INFORMATION:
; APPLICANT: Carson-Walter, Eleanor
; APPLICANT: St. Croix, Brad
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00179
; CURRENT APPLICATION NUMBER: PCT/US02/08253
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/308,829
; PRIOR FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 359
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 187
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-08253-187

Query Match      85.4%; Score 966; DB 1; Length 564;
Best Local Similarity 98.9%; Pred. No. 8.7e-89;
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFVIFSTRGTTLMKLTEDREIQROGLELQKVLPGGDTYMHGEGFERASEQIYYENRQGY 60
DB 80 MSFVIFSTRGTTLMKLTEDREIQROGLELQKVLPGGDTYMHGEGFERASEQIYYENRQGY 139
QY 61 RTASVITIALTDGELHEDLFFYSEREANRSRDGLGAIYVCVGVKDFNETQLARIADSKDHVF 120
DB 140 RTASVITIALTDGELHEDLFFYSEREANRSRDGLGAIYVCVGVKDFNETQLARIADSKDHVF 199
QY 121 PVNDGFOALQGIHHSILKSKCIEILAAEPSTICAGESFQVYVVRGNGFRHARNVDRVLCSF 180
DB 200 PVNDGFOALQGIHHSILKSKCIEILAAEPSTICAGESFQVYVVRGNGFRHARNVDRVLCSF 259
QY 181 KINDSVTLSK 190
DB 260 KINDSVTLINE 269

RESULT 8
PCT-US02-08253-232
; Sequence 232, Application PC/TUS0208253
; GENERAL INFORMATION:
; APPLICANT: Carson-Walter, Eleanor
; APPLICANT: St. Croix, Brad
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00179
; CURRENT APPLICATION NUMBER: PCT/US02/08253
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/308,829
; PRIOR FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 359
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 232
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
PCT-US02-08253-232

Query Match      85.4%; Score 966; DB 1; Length 564;
Best Local Similarity 98.9%; Pred. No. 8.7e-89;
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFVSTRTGTTLMKLTEDREQIRQGLEELQKVLPGGDTYMHGFFERASEQIYYENRQY 60
Db 80 MSFVSTRTGTTLMKLTEDREQIRQGLEELQKVLPGGDTYMHGFFERASEQIYYENRQY 139
QY 61 RTASVIALTDGELHEDLFFYSEREANRDLGAIVYCVGVKDFNETQLARIADSKDHVF 120
Db 140 RTASVIALTDGELHEDLFFYSEREANRDLGAIVYCVGVKDFNETQLARIADSKDHVF 199
QY 121 PVNDGFQALQGIHSLKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSE 180
Db 200 PVNDGFQALQGIHSLKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSE 259
QY 181 KINDSVTLK 190
Db 260 KINDSVTLNE 269

RESULT 9
US-10-038-307-20
; Sequence 20, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKATNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-20

Query Match      85.4%; Score 966; DB 6; Length 564;
Best Local Similarity 98.9%; Pred. No. 8.7e-89;
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFVSTRTGTTLMKLTEDREQIRQGLEELQKVLPGGDTYMHGFFERASEQIYYENRQY 60
Db 80 MSFVSTRTGTTLMKLTEDREQIRQGLEELQKVLPGGDTYMHGFFERASEQIYYENRQY 139
QY 61 RTASVIALTDGELHEDLFFYSEREANRDLGAIVYCVGVKDFNETQLARIADSKDHVF 120
Db 140 RTASVIALTDGELHEDLFFYSEREANRDLGAIVYCVGVKDFNETQLARIADSKDHVF 199
QY 121 PVNDGFQALQGIHSLKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSE 180
Db 200 PVNDGFQALQGIHSLKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSE 259
QY 181 KINDSVTLK 190
Db 260 KINDSVTLNE 269

RESULT 10
US-60-389-987-1823
; Sequence 1823, Application US/60389987
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M. E.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465P2
; CURRENT APPLICATION NUMBER: US/60/389,987
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 3025
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1823
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-389-987-1823

Query Match      85.4%; Score 966; DB 7; Length 564;
Best Local Similarity 98.9%; Pred. No. 8.7e-89;
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFVSTRTGTTLMKLTEDREQIRQGLEELQKVLPGGDTYMHGFFERASEQIYYENRQY 60
Db 80 MSFVSTRTGTTLMKLTEDREQIRQGLEELQKVLPGGDTYMHGFFERASEQIYYENRQY 139
QY 61 RTASVIALTDGELHEDLFFYSEREANRDLGAIVYCVGVKDFNETQLARIADSKDHVF 120
Db 140 RTASVIALTDGELHEDLFFYSEREANRDLGAIVYCVGVKDFNETQLARIADSKDHVF 199
QY 121 PVNDGFQALQGIHSLKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSE 180
Db 200 PVNDGFQALQGIHSLKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSE 259
QY 181 KINDSVTLK 190
Db 260 KINDSVTLNE 269

RESULT 11
PCT-US02-08253-194
; Sequence 194, Application PC/TUS0208253
; GENERAL INFORMATION:
; APPLICANT: Carson-Walter, Eleanor
; APPLICANT: St. Croix, Brad
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: ENDOHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00179
; CURRENT APPLICATION NUMBER: PCT/US02/08253
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/308,829
; PRIOR FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 359
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 194
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Mus musculus
PCT-US02-08253-194

Query Match      84.9%; Score 960; DB 1; Length 562;
Best Local Similarity 98.4%; Pred. No. 3.5e-88;
Matches 187; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSFVSTRTGTTLMKLTEDREQIRQGLEELQKVLPGGDTYMHGFFERASEQIYYENRQY 60
Db 78 MSFVSTRTGTTLMKLTEDREQIRQGLEELQKVLPGGDTYMHGFFERASEQIYYENRQY 137
QY 61 RTASVIALTDGELHEDLFFYSEREANRDLGAIVYCVGVKDFNETQLARIADSKDHVF 120
```



Db 138 RTASVIALTDGELHEDLFFYSEREANRSDLGAIIVYCVGKDFNETQLARIADSKDHVF 197  
QY 121 PVNDGFOALOGIIHSILKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSE 180  
Db 198 PVNDGFOALOGIIHSILKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSE 257  
QY 181 KINDSVTLK 190  
Db 258 KINDSVTLNE 267

## RESULT 12

PCT-US02-08253-301  
; Sequence 301, Application PC/TUS0208253  
; GENERAL INFORMATION:  
; APPLICANT: Carson-Walter, Eleanor  
; APPLICANT: St. Croix, Brad  
; APPLICANT: Vogelstein, Bert  
; APPLICANT: Kinzler, Kenneth  
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
; FILE REFERENCE: 1107.00179  
; CURRENT APPLICATION NUMBER: PCT/US02/08253  
; CURRENT FILING DATE: 2002-04-10  
; PRIOR APPLICATION NUMBER: 60/282,850  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/308,829  
; PRIOR FILING DATE: 2001-08-01  
; NUMBER OF SEQ ID NOS: 359  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 301  
; LENGTH: 562  
; TYPE: PRT  
; ORGANISM: Mouse  
PCT-US02-08253-301

Query Match 84.9%; Score 960; DB 1; Length 562;  
Best Local Similarity 98.4%; Pred. No. 3.5e-88;  
Matches 187; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSFVSTRTGTTLMKLTEDREQIRQGLEELQKVLPGDGYMHGFERASEQIYYENRQGY 60  
Db 78 MSFVSTRTGTTLMKLTEDREQIRQGLEELQKVLPGDGYMHGFERASEQIYYENRQGY 137  
QY 61 RTASVIALTDGELHEDLFFYSEREANRSDLGAIIVYCVGKDFNETQLARIADSKDHVF 120  
Db 138 RTASVIALTDGELHEDLFFYSEREANRSDLGAIIVYCVGKDFNETQLARIADSKDHVF 197  
QY 121 PVNDGFOALOGIIHSILKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSE 180  
Db 198 PVNDGFOALOGIIHSILKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSE 257  
QY 181 KINDSVTLK 190  
Db 258 KINDSVTLNE 267

## RESULT 13

US-10-038-307-12  
; Sequence 12, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12

; LENGTH: 534  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-038-307-12

Query Match 83.6%; Score 945; DB 6; Length 534;  
Best Local Similarity 97.9%; Pred. No. 1.1e-86;  
Matches 186; Conservative 2; Mismatches 0; Indels 2; Gaps 1;

QY 1 MSFVSTRTGTTLMKLTEDREQIRQGLEELQKVLPGDGYMHGFERASEQIYYENRQGY 60  
Db 71 MSFVSTRTGTTLMKLTEDREQIRQGLEELQKVLPGDGYMHGFERASEQIYYENRQGY 130  
QY 61 RTASVIALTDGELHEDLFFYSEREANRSDLGAIIVYCVGKDFNETQLARIADSKDHVF 120  
Db 131 RTASVIALTDGELHEDLFFYS--EANRSDLGAIIVYCVGKDFNETQLARIADSKDHVF 188  
QY 121 PVNDGFOALOGIIHSILKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSE 180  
Db 189 PVNDGFOALOGIIHSILKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSE 248  
QY 181 KINDSVTLK 190  
Db 249 KINDSVTLNE 258

## RESULT 14

US-10-038-307-10  
; Sequence 10, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 543  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-038-307-10

Query Match 83.6%; Score 945; DB 6; Length 543;  
Best Local Similarity 97.9%; Pred. No. 1.1e-86;  
Matches 186; Conservative 2; Mismatches 0; Indels 2; Gaps 1;

QY 1 MSFVSTRTGTTLMKLTEDREQIRQGLEELQKVLPGDGYMHGFERASEQIYYENRQGY 60  
Db 80 MSFVSTRTGTTLMKLTEDREQIRQGLEELQKVLPGDGYMHGFERASEQIYYENRQGY 139  
QY 61 RTASVIALTDGELHEDLFFYSEREANRSDLGAIIVYCVGKDFNETQLARIADSKDHVF 120  
Db 140 RTASVIALTDGELHEDLFFYS--EANRSDLGAIIVYCVGKDFNETQLARIADSKDHVF 197  
QY 121 PVNDGFOALOGIIHSILKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSE 180  
Db 198 PVNDGFOALOGIIHSILKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSE 257  
QY 181 KINDSVTLK 190  
Db 258 KINDSVTLNE 267

## RESULT 15

US-10-038-307-14  
; Sequence 14, Application US/10038307  
; GENERAL INFORMATION:

Fri Aug 9 10:56:54 2002

```
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-038-307-14

Query Match      83.6%; Score 945; DB 6; Length 543;
Best Local Similarity 97.9%; Pred. No. 1.1e-86;
Matches 186; Conservative 2; Mismatches 0; Indels 2; Gaps 1;

QY 1 MSFVFTSTGTTLMKLTEDREQIRQGLELELQKVLPGDYMHEGFERASEQIYYENRQGY 60
Db 80 MSFVFTSTGTTLMKLTEDREQIRQGLELELQKVLPGDYMHEGFERASEQIYYENRQGY 139
QY 61 RTASVIALTDGELHEDLFFYSEREANRDLGAIYVCVKDFNETQLARIADSKDHVF 120
Db 140 RTASVIALTDGELHEDLFFYS--EANRDLGAIYVCVKDFNETQLARIADSKDHVF 197
QY 121 PVNDGFOALQGIHSILKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVDRVLCSE 180
Db 198 PVNDGFOALQGIHSILKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVDRVLCSE 257
QY 181 KINDSVTLSK 190
Db 258 KINDSVTLINE 267
```

Search completed: August 9, 2002, 10:35:59  
Job time: 357 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:35:57 ; Search time 98.32 Seconds

(without alignments)

593.720 Million cell updates/sec

Title: us-09-970-076-8

Perfect score: 1728

Sequence: 1 MATERRALGIGFOWLSLAT.....TTHCSLHKIASGPTTAACME 333

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 528882 seqs, 175299045 residues

Total number of hits satisfying chosen parameters: 528882

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Pending\_Patents\_AA\_New:\*
- 1: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pcp.\*
  - 2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pcp.\*
  - 3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pcp.\*
  - 4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pcp.\*
  - 5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pcp.\*
  - 6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pcp.\*
  - 7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1728	100.0	333	6	US-10-038-307-2
2	1728	100.0	564	6	US-10-038-307-20
3	1718	99.4	345	6	US-10-038-307-24
4	1713.5	99.2	342	6	US-10-038-307-22
5	1650.5	95.5	551	6	US-10-038-307-18
6	1649	95.4	564	1	PCT-US02-08253-187
7	1649	95.4	564	1	PCT-US02-08253-232
8	1649	95.4	564	7	US-60-389-987-1823
9	1636	94.7	328	6	US-10-038-307-26
10	1628	94.2	543	6	US-10-038-307-14
11	1628	94.2	543	6	US-10-038-307-16
12	1618	93.6	543	6	US-10-038-307-10
13	1553	89.9	562	1	PCT-US02-08253-194
14	1553	89.9	562	1	PCT-US02-08253-301
15	1487	86.1	534	6	US-10-038-307-12
16	966	55.9	218	5	US-09-629-469A-11706
17	903	52.3	538	6	US-10-047-542-99
18	800.5	46.3	488	6	US-10-104-047-2639
19	800.5	46.3	488	6	US-10-038-307-6
20	800.5	46.3	488	7	US-60-373-595-10
21	785.5	45.5	587	6	US-10-125-540-312
22	771	44.6	487	6	US-10-038-307-8
23	669	38.7	381	6	US-10-038-307-4
24	145.5	8.4	1152	5	US-09-592-617A-43
25	138.5	8.0	1163	1	PCT-US02-10824-116
26	138.5	8.0	1163	7	US-60-369-452-33

27	134.5	7.8	646	6	US-10-155-881-22380
28	132.5	7.7	191	5	US-09-805-354-1
29	132	7.6	460	6	US-10-155-881-33525
30	131	7.6	3063	1	PCT-US02-08253-257
31	131	7.6	3063	1	PCT-US02-19669-61
32	131	7.6	3063	1	PCT-US02-19669-63
33	131	7.6	3063	6	US-10-177-293-61
34	131	7.6	3063	6	US-10-177-293-63
35	128	7.4	3051	7	US-60-360-039-5866
36	125.5	7.3	191	5	US-09-805-354-2
37	124.5	7.2	191	5	US-09-805-354-3
38	121.5	7.0	187	5	US-09-592-617A-59
39	118.5	6.9	3594	6	US-10-150-821-4
40	117.5	6.8	1207	7	US-60-389-987-1591
41	114	6.6	682	6	US-10-155-881-28757
42	114	6.6	915	5	US-09-904-920A-34
43	114	6.6	915	6	US-10-121-049-294
44	114	6.6	915	6	US-10-121-050-294
45	114	6.6	915	6	US-10-121-053-294

ALIGNMENTS

RESULT 1

US-10-038-307-2  
; Sequence 2, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 333  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-038-307-2

Query Match	100.0%	Score 1728;	DB 6;	Length 333;
Best Local Similarity	100.0%;	Pred. No. 1.7e-154;		
Matches 333;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	
Qy	1	MATERRALGIGFOWLSLATLVLCAGOGGRRDGGPCYGGFDLYFILDKSGSVLHHWN	60	
Db	1	MATERRALGIGFOWLSLATLVLCAGOGGRRDGGPCYGGFDLYFILDKSGSVLHHWN	60	
Qy	61	EIYVEQLAHFKFISPOLRMSFIVSTGTTLMKLTEDREQIROGLEELQKVLPGGDTVM	120	
Db	61	EIYVEQLAHFKFISPOLRMSFIVSTGTTLMKLTEDREQIROGLEELQKVLPGGDTVM	120	
Qy	121	HEGFERAEQIYENRQGVRTASVIALTDGELHEDLFYSERANRSDLCIAIVCYGV	180	
Db	121	HEGFERAEQIYENRQGVRTASVIALTDGELHEDLFYSERANRSDLCIAIVCYGV	180	
Qy	181	KDFNETQLARIADSKDHVPVNDGFGAOLQGIHSLIKKSCIEILAAEPSTICAGESFQVV	240	
Db	181	KDFNETQLARIADSKDHVPVNDGFGAOLQGIHSLIKKSCIEILAAEPSTICAGESFQVV	240	
Qy	241	VRNGFRHARNVDRVLCSPKINDSVTLNPKPSVEDTYLLCPAPILKEVGMKAALQVSMN	300	
Db	241	VRNGFRHARNVDRVLCSPKINDSVTLNPKPSVEDTYLLCPAPILKEVGMKAALQVSMN	300	
Qy	301	DGLSFSSSVIITTHCSLHKIASGPTTAACME	333	
Db	301	DGLSFSSSVIITTHCSLHKIASGPTTAACME	333	

Best Local Similarity 99.7%; Pred. No. 1.6e-153;  
Matches 332; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MATAERRALGIGFOWLSLATLVLCAGOGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFOWLSLATLVLCAGOGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60  
Qy 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRTTLMKLTEDREIQROGLELEOKVLPGGDTYM 120  
Db 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRTTLMKLTEDREIQROGLELEOKVLPGGDTYM 120  
Qy 121 HEGFERASEQIYYENRQGYRTASVIIALTDELHEDLFFYSEREANRSDLGAIYVCV 180  
Db 121 HEGFERASEQIYYENRQGYRTASVIIALTDELHEDLFFYSEREANRSDLGAIYVCV 180  
Qy 181 KDFNETQLARIADSKDHVPVNDGFOALOGIIHSLKSCIEILAAEPSTICAGSFQV 240  
Db 181 KDFNETQLARIADSKDHVPVNDGFOALOGIIHSLKSCIEILAAEPSTICAGSFQV 240  
Qy 241 VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300  
Qy 301 DGLSFSSSVIITTHCSLHKIASGPTTAACME 333  
Db 301 DGLSFSSSVIITTHCSLHKIASGPTTAASME 333

RESULT 4  
US-10-038-307-22  
; Sequence 22, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 22  
; LENGTH: 342  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-038-307-22

Query Match 99.2%; Score 1713.5; DB 6; Length 342;  
Best Local Similarity 97.4%; Pred. No. 4.2e-153;  
Matches 333; Conservative 0; Mismatches 0; Indels 9; Gaps 1;

Qy 1 MATAERRALGIGFOWLSLATLVLCAGOGGRRDGGP-----ACYGGFDLYFLDK 51  
Db 1 MATAERRALGIGFOWLSLATLVLCAGOGGRRDGGPMDYKDDDDKACYGGFDLYFLDK 60  
Qy 52 SGSVLHHWNEIYYFVEQLAHKFISPOLRMSFIVFSTRTTLMKLTEDREIQROGLELOK 111  
Db 61 SGSVLHHWNEIYYFVEQLAHKFISPOLRMSFIVFSTRTTLMKLTEDREIQROGLELOK 120  
Qy 112 VLPGGDTYMHGFFERASEQIYYENRQGYRTASVIIALTDELHEDLFFYSEREANRSD 171  
Db 121 VLPGGDTYMHGFFERASEQIYYENRQGYRTASVIIALTDELHEDLFFYSEREANRSD 180  
Qy 172 GAIVYCVGVKDFNETQLARIADSKDHVPVNDGFOALOGIIHSLKSCIEILAAEPSTI 231  
Db 181 GAIVYCVGVKDFNETQLARIADSKDHVPVNDGFOALOGIIHSLKSCIEILAAEPSTI 240  
Qy 232 CAGESFQVVRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGM 291  
Db 241 CAGESFQVVRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGM 300

US-10-038-307-20  
; Sequence 20, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 20  
; LENGTH: 564  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-038-307-20

Query Match 100.0%; Score 1728; DB 6; Length 564;  
Best Local Similarity 100.0%; Pred. No. 3.6e-154;  
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATAERRALGIGFOWLSLATLVLCAGOGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFOWLSLATLVLCAGOGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60  
Qy 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRTTLMKLTEDREIQROGLELEOKVLPGGDTYM 120  
Db 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRTTLMKLTEDREIQROGLELEOKVLPGGDTYM 120  
Qy 121 HEGFERASEQIYYENRQGYRTASVIIALTDELHEDLFFYSEREANRSDLGAIYVCV 180  
Db 121 HEGFERASEQIYYENRQGYRTASVIIALTDELHEDLFFYSEREANRSDLGAIYVCV 180  
Qy 181 KDFNETQLARIADSKDHVPVNDGFOALOGIIHSLKSCIEILAAEPSTICAGSFQV 240  
Db 181 KDFNETQLARIADSKDHVPVNDGFOALOGIIHSLKSCIEILAAEPSTICAGSFQV 240  
Qy 241 VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300  
Qy 301 DGLSFSSSVIITTHCSLHKIASGPTTAACME 333  
Db 301 DGLSFSSSVIITTHCSLHKIASGPTTAACME 333

RESULT 3  
US-10-038-307-24  
; Sequence 24, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 24  
; LENGTH: 345  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-038-307-24

Query Match 99.4%; Score 1718; DB 6; Length 345;



Fri Aug 9 10:57:08 2002

121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSEREANRSDLGAIYVCV 180  
Db 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSEREANRSDLGAIYVCV 180  
181 KDFNETQLARIADSKDHVPVNDGFGFQALQGIHSLKSKSCIEILAAEPSTICAGESFQV 240  
Db 181 KDFNETQLARIADSKDHVPVNDGFGFQALQGIHSLKSKSCIEILAAEPSTICAGESFQV 240  
241 VRNGGFRHARNVDRVLCSEKINDSVTLNEKPFSEVDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRNGGFRHARNVDRVLCSEKINDSVTLNEKPFSEVDTYLLCPAPILKEVGMKAALQVSMN 300  
301 DGLSFSSSVIIITTHCS 318  
Db 301 DGLSFSSSVIIITTHCS 318  
Query Match 95.4%; Score 1649; DB 7; Length 564;  
Best Local Similarity 100.0%; Pred. No. 9.9e-147;  
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 MATAERRALGIGFQWLSLATLVLCAGQGRRDGGPACYGFGDLYFLDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFQWLSLATLVLCAGQGRRDGGPACYGFGDLYFLDKSGSVLHHWN 60  
61 EIYFVEQLAHKFISPOLRMSFVSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Db 61 EIYFVEQLAHKFISPOLRMSFVSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSEREANRSDLGAIYVCV 180  
Db 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSEREANRSDLGAIYVCV 180  
181 KDFNETQLARIADSKDHVPVNDGFGFQALQGIHSLKSKSCIEILAAEPSTICAGESFQV 240  
Db 181 KDFNETQLARIADSKDHVPVNDGFGFQALQGIHSLKSKSCIEILAAEPSTICAGESFQV 240  
241 VRNGGFRHARNVDRVLCSEKINDSVTLNEKPFSEVDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRNGGFRHARNVDRVLCSEKINDSVTLNEKPFSEVDTYLLCPAPILKEVGMKAALQVSMN 300  
301 DGLSFSSSVIIITTHCS 318  
Db 301 DGLSFSSSVIIITTHCS 318  
Query Match 94.7%; Score 1636; DB 6; Length 328;  
Best Local Similarity 100.0%; Pred. No. 7.8e-146;  
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 MATAERRALGIGFQWLSLATLVLCAGQGRRDGGPACYGFGDLYFLDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFQWLSLATLVLCAGQGRRDGGPACYGFGDLYFLDKSGSVLHHWN 60  
61 EIYFVEQLAHKFISPOLRMSFVSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Db 61 EIYFVEQLAHKFISPOLRMSFVSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSEREANRSDLGAIYVCV 180  
Db 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSEREANRSDLGAIYVCV 180  
181 KDFNETQLARIADSKDHVPVNDGFGFQALQGIHSLKSKSCIEILAAEPSTICAGESFQV 240  
Db 181 KDFNETQLARIADSKDHVPVNDGFGFQALQGIHSLKSKSCIEILAAEPSTICAGESFQV 240  
241 VRNGGFRHARNVDRVLCSEKINDSVTLNEKPFSEVDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRNGGFRHARNVDRVLCSEKINDSVTLNEKPFSEVDTYLLCPAPILKEVGMKAALQVSMN 300  
301 DGLSFSSSVIIITTHCS 318  
Db 301 DGLSFSSSVIIITTHCS 318  
Query Match 94.2%; Score 1628; DB 6; Length 543;  
Best Local Similarity 99.4%; Pred. No. 9e-145;  
Matches 316; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

Sequence 26, Application US/10038307  
GENERAL INFORMATION:  
APPLICANT: James B. ROTTMAN  
APPLICANT: Theresa L. O'KEEFE  
APPLICANT: Engin OZKAYNAK  
APPLICANT: Judith J. HEALEY  
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
FILE REFERENCE: 7853-253-999  
CURRENT APPLICATION NUMBER: US/10/038,307  
CURRENT FILING DATE: 2002-06-28  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 26  
LENGTH: 328  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-038-307-26  
Query Match 94.7%; Score 1636; DB 6; Length 328;  
Best Local Similarity 100.0%; Pred. No. 7.8e-146;  
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 MATAERRALGIGFQWLSLATLVLCAGQGRRDGGPACYGFGDLYFLDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFQWLSLATLVLCAGQGRRDGGPACYGFGDLYFLDKSGSVLHHWN 60  
61 EIYFVEQLAHKFISPOLRMSFVSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Db 61 EIYFVEQLAHKFISPOLRMSFVSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSEREANRSDLGAIYVCV 180  
Db 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSEREANRSDLGAIYVCV 180  
181 KDFNETQLARIADSKDHVPVNDGFGFQALQGIHSLKSKSCIEILAAEPSTICAGESFQV 240  
Db 181 KDFNETQLARIADSKDHVPVNDGFGFQALQGIHSLKSKSCIEILAAEPSTICAGESFQV 240  
241 VRNGGFRHARNVDRVLCSEKINDSVTLNEKPFSEVDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRNGGFRHARNVDRVLCSEKINDSVTLNEKPFSEVDTYLLCPAPILKEVGMKAALQVSMN 300  
301 DGLSFSSSVIIITTHCS 316  
Db 301 DGLSFSSSVIIITTHCS 316  
RESULT 10  
US-10-038-307-14  
Sequence 14, Application US/10038307  
GENERAL INFORMATION:  
APPLICANT: James B. ROTTMAN  
APPLICANT: Theresa L. O'KEEFE  
APPLICANT: Engin OZKAYNAK  
APPLICANT: Judith J. HEALEY  
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
FILE REFERENCE: 7853-253-999  
CURRENT APPLICATION NUMBER: US/10/038,307  
CURRENT FILING DATE: 2002-06-28  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 14  
LENGTH: 543  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-038-307-14  
Query Match 94.2%; Score 1628; DB 6; Length 543;  
Best Local Similarity 99.4%; Pred. No. 9e-145;  
Matches 316; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

RESULT 9  
US-10-038-307-26

```
QY 1 MATAERRALGIGFOWLSLATLVLCAGQGRRDGGPACYGDFLYFILDKSGSVLHHWN 60
Db 1 MATAERRALGIGFOWLSLATLVLCAGQGRRDGGPACYGDFLYFILDKSGSVLHHWN 60
QY 61 EIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120
Db 61 EIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120
QY 121 HEGFERASEQIYYENRQGYRTASVIITLTDGELHEDLFFYSREANRSDLGAIYVCVG 180
Db 121 HEGFERASEQIYYENRQGYRTASVIITLTDGELHEDLFFYS--EANRSDLGAIYVCVG 178
QY 181 KDFNETQIARIADSKDHVPVNDGFOALQGIHSILKKSCIEILAAEPSTICAGESFOVV 240
Db 179 KDFNETQIARIADSKDHVPVNDGFOALQGIHSILKKSCIEILAAEPSTICAGESFOVV 238
QY 241 VRNGFRHARNVDRVLCSEFKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300
Db 239 VRNGFRHARNVDRVLCSEFKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 298
QY 301 DGLSFISSSVIITTHCS 318
Db 299 DGLSFISSSVIITTHCS 316

RESULT 11
US-10-038-307-16
; Sequence 16, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038, 307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-16

Query Match 94.2%; Score 1628; DB 6; Length 543;
Best Local Similarity 99.4%; Pred. No. 9e-145;
Matches 316; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 MATAERRALGIGFOWLSLATLVLCAGQGRRDGGPACYGDFLYFILDKSGSVLHHWN 60
Db 1 MATAERRALGIGFOWLSLATLVLCAGQGRRDGGPACYGDFLYFILDKSGSVLHHWN 60
QY 61 EIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120
Db 61 EIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120
QY 121 HEGFERASEQIYYENRQGYRTASVIITLTDGELHEDLFFYSREANRSDLGAIYVCVG 180
Db 121 HEGFERASEQIYYENRQGYRTASVIITLTDGELHEDLFFYS--EANRSDLGAIYVCVG 178
QY 181 KDFNETQIARIADSKDHVPVNDGFOALQGIHSILKKSCIEILAAEPSTICAGESFOVV 240
Db 179 KDFNETQIARIADSKDHVPVNDGFOALQGIHSILKKSCIEILAAEPSTICAGESFOVV 238
QY 241 VRNGFRHARNVDRVLCSEFKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300
Db 239 VRNGFRHARNVDRVLCSEFKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 298
QY 301 DGLSFISSSVIITTHCS 318
Db 299 DGLSFISSSVIITTHCS 316
```

```
RESULT 12
US-10-038-307-10
; Sequence 10, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038, 307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-10

Query Match 93.6%; Score 1618; DB 6; Length 543;
Best Local Similarity 99.1%; Pred. No. 7.8e-144;
Matches 315; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

QY 1 MATAERRALGIGFOWLSLATLVLCAGQGRRDGGPACYGDFLYFILDKSGSVLHHWN 60
Db 1 MATAERRALGIGFOWLSLATLVLCAGQGRRDGGPACYGDFLYFILDKSGSVLHHWN 60
QY 61 EIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120
Db 61 EIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120
QY 121 HEGFERASEQIYYENRQGYRTASVIITLTDGELHEDLFFYSREANRSDLGAIYVCVG 180
Db 121 HEGFERASEQIYYENRQGYRTASVIITLTDGELHEDLFFYS--EANRSDLGAIYVCVG 178
QY 181 KDFNETQIARIADSKDHVPVNDGFOALQGIHSILKKSCIEILAAEPSTICAGESFOVV 240
Db 179 KDFNETQIARIADSKDHVPVNDGFOALQGIHSILKKSCIEILAAEPSTICAGESFOVV 238
QY 241 VRNGFRHARNVDRVLCSEFKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300
Db 239 VRNGFRHARNVDRVLCSEFKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 298
QY 301 DGLSFISSSVIITTHCS 318
Db 299 DGLSFISSSVIITTHCS 316

RESULT 13
PCT-US02-08253-194
; Sequence 194, Application PC/TUS0208253
; GENERAL INFORMATION:
; APPLICANT: Carson-Walter, Eleanor
; APPLICANT: St. Croix, Brad
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00179
; CURRENT APPLICATION NUMBER: PCT/US02/08253
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/308,829
; PRIOR FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 359
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 194
; LENGTH: 562
; TYPE: PRT
```

ORGANISM: Mus musculus  
PCT-US02-08253-194

Query Match 89.9%; Score 1553; DB 1; Length 562;  
Best Local Similarity 96.1%; Pred. No. 1.1e-137;  
Matches 298; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 9 LGIGFQWLSLATLVLCAGOGGREDGGPACYGDFLYFILDKSGSVLHHWNEIYFVEQ 68  
Db 7 LGAGLGLCAALVLCAGHGGREDGGPACYGDFLYFILDKSGSVLHHWNEIYFVEQ 66

Qy 69 LAHKFISPOLRMSFIVFSTRTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFERAS 128  
Db 67 LAHREISPOLRMSFIVFSTRTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFERAS 126

Qy 129 EQIYYENRQGYRTASVIITALDGHEDLFFYSEREANRSDLGAIYVCVKDFNETQL 188  
Db 127 EQIYYENRQGYRTASVIITALDGHEDLFFYSEREANRSDLGAIYVCVKDFNETQL 186

Qy 189 ARIADSKDHVPVNDGFOALOGIIHSILKSCIEILAAEPSTICAGESFQVVRNGNGFRH 248  
Db 187 ARIADSKDHVPVNDGFOALOGIIHSILKSCIEILAAEPSTICAGESFQVVRNGNGFRH 246

Qy 249 ARNDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMNDGLSFSS 308  
Db 247 ARNDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMNDGLSFSS 306

Qy 309 SVIITTHCS 318  
Db 307 SVIITTHCS 316

RESULT 14  
PCT-US02-08253-301  
; Sequence 301, Application PC/TUS0208253  
; GENERAL INFORMATION:  
; APPLICANT: Carson-Walter, Eleanor  
; APPLICANT: St. Croix, Brad  
; APPLICANT: Vogelstein, Bert  
; APPLICANT: Kinzler, Kenneth  
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
; FILE REFERENCE: 1107.00179  
; CURRENT APPLICATION NUMBER: PCT/US02/08253  
; CURRENT FILING DATE: 2002-04-10  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR FILING DATE: 2001-08-01  
; NUMBER OF SEQ ID NOS: 359  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 301  
; LENGTH: 562  
; TYPE: PRT  
; ORGANISM: Mouse  
PCT-US02-08253-301

Query Match 89.9%; Score 1553; DB 1; Length 562;  
Best Local Similarity 96.1%; Pred. No. 1.1e-137;  
Matches 298; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 9 LGIGFQWLSLATLVLCAGOGGREDGGPACYGDFLYFILDKSGSVLHHWNEIYFVEQ 68  
Db 7 LGAGLGLCAALVLCAGHGGREDGGPACYGDFLYFILDKSGSVLHHWNEIYFVEQ 66

Qy 69 LAHKFISPOLRMSFIVFSTRTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFERAS 128  
Db 67 LAHREISPOLRMSFIVFSTRTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFERAS 126

Qy 129 EQIYYENRQGYRTASVIITALDGHEDLFFYSEREANRSDLGAIYVCVKDFNETQL 188  
Db 127 EQIYYENRQGYRTASVIITALDGHEDLFFYSEREANRSDLGAIYVCVKDFNETQL 186

Qy 189 ARIADSKDHVPVNDGFOALOGIIHSILKSCIEILAAEPSTICAGESFQVVRNGNGFRH 248  
Db 187 ARIADSKDHVPVNDGFOALOGIIHSILKSCIEILAAEPSTICAGESFQVVRNGNGFRH 246

Qy 249 ARNDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMNDGLSFSS 308  
Db 247 ARNDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMNDGLSFSS 306

Qy 309 SVIITTHCS 318  
Db 307 SVIITTHCS 316

RESULT 15  
US-10-038-307-12  
; Sequence 12, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038, 307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 534  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-038-307-12

Query Match 86.1%; Score 1487; DB 6; Length 534;  
Best Local Similarity 94.5%; Pred. No. 1.7e-131;  
Matches 293; Conservative 3; Mismatches 8; Indels 6; Gaps 2;

Qy 9 LGIGFQWLSLATLVLCAGOGGREDGGPACYGDFLYFILDKSGSVLHHWNEIYFVEQ 68  
Db 4 LGLLFCLVTLPCVLT---SGRRREDGGPACYGDFLYFILDKSGSVLHHWNEIYFVEQ 59

Qy 69 LAHKFISPOLRMSFIVFSTRTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFERAS 128  
Db 60 LAHKFISPOLRMSFIVFSTRTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFERAS 119

Qy 129 EQIYYENRQGYRTASVIITALDGHEDLFFYSEREANRSDLGAIYVCVKDFNETQL 188  
Db 120 EQIYYENRQGYRTASVIITALDGHEDLFFYS--EANSRDLGAIYVCVKDFNETQL 177

Qy 189 ARIADSKDHVPVNDGFOALOGIIHSILKSCIEILAAEPSTICAGESFQVVRNGNGFRH 248  
Db 178 ARIADSKDHVPVNDGFOALOGIIHSILKSCIEILAAEPSTICAGESFQVVRNGNGFRH 237

Qy 249 ARNDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMNDGLSFSS 308  
Db 238 ARNDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMNDGLSFSS 297

Qy 309 SVIITTHCS 318  
Db 298 SVIITTHCS 307

Search completed: August 9, 2002, 10:35:58  
Job time: 356 sec





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: August 9, 2002, 10:44:31 ; Search time 507.33 seconds  
(without alignments)  
231.032 Million cell updates/sec

Title: US-09-970-076-8

Perfect score: 1728

Sequence: 1 MATAERRALGIGFOWLSLAT.....TTHCSLKHASGPTTAACME 333

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Pending\_Patents\_AA\_Main:\*
- 1: /cgn2\_6/ptodata/2/paa/PCUTUS\_COMB.pcp.\*
  - 2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pcp.\*
  - 3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pcp.\*
  - 4: /cgn2\_6/ptodata/2/paa/US08\_COMB.pcp.\*
  - 5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pcp.\*
  - 6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pcp.\*
  - 7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pcp.\*
  - 8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pcp.\*
  - 9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pcp.\*
  - 10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pcp.\*
  - 11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pcp.\*
  - 12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pcp.\*
  - 13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pcp.\*
  - 14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pcp.\*
  - 15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pcp.\*
  - 16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pcp.\*
  - 17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pcp.\*
  - 18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pcp.\*
  - 19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pcp.\*
  - 20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pcp.\*
  - 21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pcp.\*
  - 22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pcp.\*
  - 23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pcp.\*
  - 24: /cgn2\_6/ptodata/2/paa/US100\_COMB.pcp.\*
  - 25: /cgn2\_6/ptodata/2/paa/US101\_COMB.pcp.\*
  - 26: /cgn2\_6/ptodata/2/paa/US60\_COMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1728	100.0	333	1	PCT-US99-31025-9
2	1728	100.0	333	16	US-09-223-546-9
3	1728	100.0	333	18	US-09-471-179-9
4	1728	100.0	333	19	US-09-599-596-9
5	1728	100.0	333	21	US-09-796-753-12
6	1728	100.0	333	23	US-09-970-076-8
7	1649	95.4	368	23	US-09-970-076-2

8	1649	95.4	403	1	PCT-US00-30045-94	Sequence 94, Appl
9	1649	95.4	403	1	PCT-US01-11988-621	Sequence 621, App
10	1649	95.4	403	22	US-09-833-245-621	Sequence 621, App
11	1649	95.4	564	23	US-09-918-175-187	Sequence 187, App
12	1649	95.4	564	23	US-09-918-175-232	Sequence 232, App
13	1649	95.4	564	23	US-09-970-076-6	Sequence 6, Appl
14	1644	95.1	403	1	PCT-US00-30045-125	Sequence 125, App
15	1644	95.1	403	1	PCT-US01-11988-620	Sequence 620, App
16	1644	95.1	403	22	US-09-833-245-620	Sequence 620, App
17	1593	92.2	306	1	PCT-US99-31025-30	Sequence 30, Appl
18	1593	92.2	306	1	PCT-US99-31025-51	Sequence 51, Appl
19	1593	92.2	306	18	US-09-471-179-30	Sequence 30, Appl
20	1593	92.2	306	18	US-09-471-179-51	Sequence 51, Appl
21	1565	90.6	301	1	PCT-US99-31025-132	Sequence 132, App
22	1565	90.6	301	18	US-09-471-179-132	Sequence 132, App
23	1553	89.9	562	23	US-09-918-175-194	Sequence 194, App
24	1553	89.9	562	23	US-09-918-175-301	Sequence 301, App
25	1430	82.8	274	1	PCT-US99-31025-50	Sequence 50, Appl
26	1430	82.8	274	18	US-09-471-179-50	Sequence 50, Appl
27	1392	80.6	297	18	US-09-488-725A-2121	Sequence 2121, Ap
28	966	55.9	218	23	US-09-970-076-10	Sequence 10, Appl
29	895	51.8	172	1	PCT-US99-31025-105	Sequence 105, App
30	895	51.8	172	18	US-09-471-179-105	Sequence 105, App
31	801.5	46.4	488	1	PCT-US00-05226-59	Sequence 59, Appl
32	801.5	46.4	488	19	US-09-516-745-59	Sequence 59, Appl
33	800.5	46.3	487	1	PCT-US00-05226-107	Sequence 107, App
34	800.5	46.3	487	19	US-09-516-745-107	Sequence 107, App
35	800.5	46.3	488	1	PCT-US00-05226-2	Sequence 2, Appl
36	800.5	46.3	488	19	US-09-516-745-2	Sequence 2, Appl
37	800.5	46.3	488	21	US-09-796-753-52	Sequence 52, Appl
38	800.5	46.3	488	26	US-60-318-891-10	Sequence 10, Appl
39	800.5	46.3	488	26	US-60-318-905-20	Sequence 20, Appl
40	800.5	46.3	488	26	US-60-322-468-6	Sequence 6, Appl
41	800.5	46.3	488	26	US-60-322-732-10	Sequence 10, Appl
42	800.5	46.3	488	26	US-60-322-790-16	Sequence 6, Appl
43	800.5	46.3	488	26	US-60-323-078-22	Sequence 22, Appl
44	800.5	46.3	488	26	US-60-324-050-2	Sequence 2, Appl
45	800.5	46.3	488	26	US-60-324-050-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1  
PCT-US99-31025-9  
; Sequence 9, Application PC/TUS9931025  
; GENERAL INFORMATION:  
; APPLICANT: Millennium Pharmaceuticals, Inc.  
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS  
; TITLE OF INVENTION: ENCODING THEM  
; FILE REFERENCE: 7853-173-228  
; CURRENT APPLICATION NUMBER: PCT/US99/31025  
; CURRENT FILING DATE: 1999-12-23  
; EARLIER APPLICATION NUMBER: 09/223,546  
; EARLIER FILING DATE: 1998-12-30  
; NUMBER OF SEQ ID NOS: 135  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 333  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US99-31025-9

Query Match 100.0%; Score 1728; DB 1; Length 333;  
Best Local Similarity 100.0%; Pred. No. 5.2e-168;  
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MATAERRALGIGFOWLSLATVLICAGOCGRREDGCPACYGFDLYFLDKSGSVLHHN 60  
DB 1 MATAERRALGIGFOWLSLATVLICAGOCGRREDGCPACYGFDLYFLDKSGSVLHHN 60  
QY 61 EYIFVEQLAHKFIISPOLRMSFVSTRTTLMLKLTEDREQIRQGLEELQKVLPGDVTM 120

Fri Aug 9 10:57:07 2002

FILE REFERENCE: 7853-173  
CURRENT APPLICATION NUMBER: US/09/471,179  
CURRENT FILING DATE: 1999-12-23  
PRIOR APPLICATION NUMBER: 09/223,546  
PRIOR FILING DATE: 1998-12-30  
NUMBER OF SEQ ID NOS: 135  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9  
LENGTH: 333  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-471-179-9

Query Match 100.0%; Score 1728; DB 18; Length 333;  
Best Local Similarity 100.0%; Pred. No. 5.2e-168;  
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATAERRALGIGFOWLSLATLVLCAGGGRREDGGPACYGDFLYFILDKSGSVLHHWN 60  
DB 1 MATAERRALGIGFOWLSLATLVLCAGGGRREDGGPACYGDFLYFILDKSGSVLHHWN 60  
QY 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREIQRLGELQKVLPGGDTYM 120  
DB 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREIQRLGELQKVLPGGDTYM 120  
QY 121 HGFERASEQIYYENRQGYRTASVIAITDGHEDLFFYSERANRSDLGAIYVCVGV 180  
DB 121 HGFERASEQIYYENRQGYRTASVIAITDGHEDLFFYSERANRSDLGAIYVCVGV 180  
QY 181 KDFNETQLARIADSKDHPVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQV 240  
DB 181 KDFNETQLARIADSKDHPVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQV 240  
QY 241 VRNGFRHARNVDRLCSFKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
DB 241 VRNGFRHARNVDRLCSFKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
QY 301 DGLSFISSSVITTTCHSLHKIASGPTTAACME 333  
DB 301 DGLSFISSSVITTTCHSLHKIASGPTTAACME 333

RESULT 4  
US-09-599-596-9  
Sequence 9, Application US/09599596  
GENERAL INFORMATION:  
APPLICANT: Holtzman, Douglas  
TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM  
FILE REFERENCE: 09404/066001  
CURRENT APPLICATION NUMBER: US/09/599,596  
CURRENT FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: 09/223,546  
PRIOR FILING DATE: 1998-12-30  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 9  
LENGTH: 333  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-599-596-9

Query Match 100.0%; Score 1728; DB 19; Length 333;  
Best Local Similarity 100.0%; Pred. No. 5.2e-168;  
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATAERRALGIGFOWLSLATLVLCAGGGRREDGGPACYGDFLYFILDKSGSVLHHWN 60  
DB 1 MATAERRALGIGFOWLSLATLVLCAGGGRREDGGPACYGDFLYFILDKSGSVLHHWN 60  
QY 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREIQRLGELQKVLPGGDTYM 120

DB 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREIQRLGELQKVLPGGDTYM 120  
QY 121 HGFERASEQIYYENRQGYRTASVIAITDGHEDLFFYSERANRSDLGAIYVCVGV 180  
DB 121 HGFERASEQIYYENRQGYRTASVIAITDGHEDLFFYSERANRSDLGAIYVCVGV 180  
QY 181 KDFNETQLARIADSKDHPVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQV 240  
DB 181 KDFNETQLARIADSKDHPVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQV 240  
QY 241 VRNGFRHARNVDRLCSFKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
DB 241 VRNGFRHARNVDRLCSFKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
QY 301 DGLSFISSSVITTTCHSLHKIASGPTTAACME 333  
DB 301 DGLSFISSSVITTTCHSLHKIASGPTTAACME 333

RESULT 2  
US-09-223-546-9  
Sequence 9, Application US/09223546  
GENERAL INFORMATION:  
APPLICANT: Holtzman, Douglas  
TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM  
FILE REFERENCE: 09404/066001  
CURRENT APPLICATION NUMBER: US/09/223,546  
CURRENT FILING DATE: 1998-12-30  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 9  
LENGTH: 333  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-223-546-9

Query Match 100.0%; Score 1728; DB 16; Length 333;  
Best Local Similarity 100.0%; Pred. No. 5.2e-168;  
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATAERRALGIGFOWLSLATLVLCAGGGRREDGGPACYGDFLYFILDKSGSVLHHWN 60  
DB 1 MATAERRALGIGFOWLSLATLVLCAGGGRREDGGPACYGDFLYFILDKSGSVLHHWN 60  
QY 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREIQRLGELQKVLPGGDTYM 120  
DB 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREIQRLGELQKVLPGGDTYM 120  
QY 121 HGFERASEQIYYENRQGYRTASVIAITDGHEDLFFYSERANRSDLGAIYVCVGV 180  
DB 121 HGFERASEQIYYENRQGYRTASVIAITDGHEDLFFYSERANRSDLGAIYVCVGV 180  
QY 181 KDFNETQLARIADSKDHPVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQV 240  
DB 181 KDFNETQLARIADSKDHPVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQV 240  
QY 241 VRNGFRHARNVDRLCSFKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
DB 241 VRNGFRHARNVDRLCSFKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
QY 301 DGLSFISSSVITTTCHSLHKIASGPTTAACME 333  
DB 301 DGLSFISSSVITTTCHSLHKIASGPTTAACME 333

RESULT 3  
US-09-471-179-9  
Sequence 9, Application US/09471179  
GENERAL INFORMATION:  
APPLICANT: Holtzman, Douglas  
TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM

Db 61 EIYFEVQLAHKFISPOLRMSFIVFSTGRITLMKLTEDREQIRGLEELQKVLPGGDTYM 120  
QY 121 HEGFERASEQIYYENRQGYRTASVIAITDGLHEDLFFYSEREANRSDLGAIYVCV 180  
Db 121 HEGFERASEQIYYENRQGYRTASVIAITDGLHEDLFFYSEREANRSDLGAIYVCV 180  
QY 181 KDFNETQIARIADSKDHVPVNDGFOALQGIHSLKKSCIEILAAEPSTICAGESFQV 240  
Db 181 KDFNETQIARIADSKDHVPVNDGFOALQGIHSLKKSCIEILAAEPSTICAGESFQV 240  
QY 241 VRGNGFRHARNVDRLVCLSFKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRGNGFRHARNVDRLVCLSFKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300  
QY 301 DGLSFSSSVIITTHCSLHKIASGPTTAAACME 333  
Db 301 DGLSFSSSVIITTHCSLHKIASGPTTAAACME 333

RESULT 5  
US-09-796-753-12  
; Sequence 12, Application US/09796753  
; GENERAL INFORMATION:  
; APPLICANT: McCarthy, Sean A.  
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF  
; FILE REFERENCE: 7853-227-999  
; CURRENT APPLICATION NUMBER: US/09796,753  
; CURRENT FILING DATE: 2001-03-01  
; PRIOR APPLICATION NUMBER: 09/183,175  
; PRIOR FILING DATE: 1998-10-30  
; PRIOR APPLICATION NUMBER: 09/223,094  
; PRIOR FILING DATE: 1998-12-30  
; PRIOR APPLICATION NUMBER: 09/223,546  
; PRIOR FILING DATE: 1998-12-30  
; PRIOR APPLICATION NUMBER: 09/224,246  
; PRIOR FILING DATE: 1998-12-30  
; PRIOR APPLICATION NUMBER: 09/259,388  
; PRIOR FILING DATE: 1999-02-26  
; PRIOR APPLICATION NUMBER: 60/122,458  
; PRIOR FILING DATE: 1999-03-01  
; PRIOR APPLICATION NUMBER: 09/312,359  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 09/336,536  
; PRIOR FILING DATE: 1999-06-18  
; PRIOR APPLICATION NUMBER: 09/342,687  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: 09/345,464  
; PRIOR FILING DATE: 1999-06-30  
; PRIOR APPLICATION NUMBER: 09/365,164  
; PRIOR FILING DATE: 1999-07-30  
; PRIOR APPLICATION NUMBER: 09/399,723  
; PRIOR FILING DATE: 1999-09-20  
; PRIOR APPLICATION NUMBER: 09/409,534  
; PRIOR FILING DATE: 1999-09-30  
; PRIOR APPLICATION NUMBER: 09/471,179  
; PRIOR FILING DATE: 1999-12-23  
; PRIOR APPLICATION NUMBER: 09/474,071  
; PRIOR FILING DATE: 1999-12-29  
; PRIOR APPLICATION NUMBER: 09/474,072  
; PRIOR FILING DATE: 1999-12-29  
; PRIOR APPLICATION NUMBER: 09/514,010  
; PRIOR FILING DATE: 2000-02-25  
; PRIOR APPLICATION NUMBER: 09/516,745  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 09/572,002  
; PRIOR FILING DATE: 2000-05-14  
; PRIOR APPLICATION NUMBER: 09/597,993  
; PRIOR FILING DATE: 2000-06-19  
; PRIOR APPLICATION NUMBER: 09/599,596  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 09/630,334  
; PRIOR FILING DATE: 2000-07-31  
; PRIOR APPLICATION NUMBER: 09/606,565

; PRIOR FILING DATE: 2000-06-29  
; PRIOR APPLICATION NUMBER: 09/606,317  
; PRIOR FILING DATE: 2000-06-29  
; PRIOR APPLICATION NUMBER: 09/665,666  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: 09/677,751  
; PRIOR FILING DATE: 2000-09-30  
; NUMBER OF SEQ ID NOS: 162  
; SEQ ID NO 12  
; LENGTH: 333  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-796-753-12  
  
Query Match 100.0%; Score 1728; DB 21; Length 333;  
Best Local Similarity 100.0%; Pred. No. 5.2e-188;  
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MATAERRALGIGFOWLSLATLVLCAGQGRREDGGPACYGCGFDLYFILDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFOWLSLATLVLCAGQGRREDGGPACYGCGFDLYFILDKSGSVLHHWN 60  
QY 61 EIYFEVQLAHKFISPOLRMSFIVFSTGRITLMKLTEDREQIRGLEELQKVLPGGDTYM 120  
Db 61 EIYFEVQLAHKFISPOLRMSFIVFSTGRITLMKLTEDREQIRGLEELQKVLPGGDTYM 120  
QY 121 HEGFERASEQIYYENRQGYRTASVIAITDGLHEDLFFYSEREANRSDLGAIYVCV 180  
Db 121 HEGFERASEQIYYENRQGYRTASVIAITDGLHEDLFFYSEREANRSDLGAIYVCV 180  
QY 181 KDFNETQIARIADSKDHVPVNDGFOALQGIHSLKKSCIEILAAEPSTICAGESFQV 240  
Db 181 KDFNETQIARIADSKDHVPVNDGFOALQGIHSLKKSCIEILAAEPSTICAGESFQV 240  
QY 241 VRGNGFRHARNVDRLVCLSFKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRGNGFRHARNVDRLVCLSFKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300  
QY 301 DGLSFSSSVIITTHCSLHKIASGPTTAAACME 333  
Db 301 DGLSFSSSVIITTHCSLHKIASGPTTAAACME 333  
  
RESULT 6  
US-09-970-076-8  
; Sequence 8, Application US/09970076  
; GENERAL INFORMATION:  
; APPLICANT: Young, John A.T.  
; APPLICANT: Bradley, Kenneth A.  
; APPLICANT: Collier, Robert J.  
; APPLICANT: Mogridge, Jeremy S.  
; TITLE OF INVENTION: Anthrax Toxin Receptor  
; FILE REFERENCE: 960296.97745  
; CURRENT APPLICATION NUMBER: US/09/970,076  
; CURRENT FILING DATE: 2001-10-03  
; PRIOR APPLICATION NUMBER: 60/251,481  
; PRIOR FILING DATE: 2000-12-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 333  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-970-076-8  
  
Query Match 100.0%; Score 1728; DB 23; Length 333;  
Best Local Similarity 100.0%; Pred. No. 5.2e-168;  
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MATAERRALGIGFOWLSLATLVLCAGQGRREDGGPACYGCGFDLYFILDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFOWLSLATLVLCAGQGRREDGGPACYGCGFDLYFILDKSGSVLHHWN 60

Fri Aug 9 10:57:07 2002

Db 1 MATARRALGIGFQWLSLATLVLCAGQGRRDGGPACYGFGFDLYFILDKSGSVLHHWN 60  
Qy 61 EIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Db 61 EIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Qy 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSERANRSDLGAIYVCVG 180  
Db 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSERANRSDLGAIYVCVG 180  
Qy 181 KDFNETOLARIADSKDHVPVNDGFOALQGIHLSILKKSCEIILAAEPSTICAGESFOVV 240  
Db 181 KDFNETOLARIADSKDHVPVNDGFOALQGIHLSILKKSCEIILAAEPSTICAGESFOVV 240  
Qy 241 VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Qy 301 DGLSFSSSVIITTHCS 318  
Db 301 DGLSFSSSVIITTHCS 318  
Query Match 7 95.4%; Score 1649; DB 23; Length 368;  
Best Local Similarity 100.0%; Pred. No. 7.7e-160; Indels 0; Gaps 0;  
Matches 318; Conservative 0; Mismatches 0;  
GENERAL INFORMATION:  
APPLICANT: Young, John A.T.  
APPLICANT: Bradley, Kenneth A.  
APPLICANT: Collier, Robert J.  
APPLICANT: Mogridge, Jeremy S.  
FILE OF INVENTION: Anthrax Toxin Receptor  
FILE REFERENCE: 960296.97745  
CURRENT APPLICATION NUMBER: US/09/970,076  
CURRENT FILING DATE: 2001-10-03  
PRIOR APPLICATION NUMBER: 60/251,481  
PRIOR FILING DATE: 2000-12-05  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 2  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-970-076-2

Qy 1 MATARRALGIGFQWLSLATLVLCAGQGRRDGGPACYGFGFDLYFILDKSGSVLHHWN 60  
Db 1 MATARRALGIGFQWLSLATLVLCAGQGRRDGGPACYGFGFDLYFILDKSGSVLHHWN 60  
Qy 61 EIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Db 61 EIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Qy 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSERANRSDLGAIYVCVG 180  
Db 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSERANRSDLGAIYVCVG 180  
Qy 181 KDFNETOLARIADSKDHVPVNDGFOALQGIHLSILKKSCEIILAAEPSTICAGESFOVV 240  
Db 181 KDFNETOLARIADSKDHVPVNDGFOALQGIHLSILKKSCEIILAAEPSTICAGESFOVV 240  
Qy 241 VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Qy 301 DGLSFSSSVIITTHCS 318  
Db 301 DGLSFSSSVIITTHCS 318

RESULT 8  
PCT-US00-30045-94  
; Sequence 94, Application PC/TUS0030045  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: 28 Human Secreted Proteins  
; FILE REFERENCE: PS708PCT  
; CURRENT APPLICATION NUMBER: PCT/US00/30045  
; CURRENT FILING DATE: 2000-11-01  
; PRIOR APPLICATION NUMBER: 60/163,581  
; PRIOR FILING DATE: 1999-11-05  
; PRIOR APPLICATION NUMBER: 60/215,133  
; PRIOR FILING DATE: 2000-06-30  
; NUMBER OF SEQ ID NOS: 201  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 94  
; LENGTH: 403  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US00-30045-94  
Query Match 95.4%; Score 1649; DB 1; Length 403;  
Best Local Similarity 100.0%; Pred. No. 8.9e-160; Indels 0; Gaps 0;  
Matches 318; Conservative 0; Mismatches 0;  
Qy 1 MATARRALGIGFQWLSLATLVLCAGQGRRDGGPACYGFGFDLYFILDKSGSVLHHWN 60  
Db 1 MATARRALGIGFQWLSLATLVLCAGQGRRDGGPACYGFGFDLYFILDKSGSVLHHWN 60  
Qy 61 EIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Db 61 EIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Qy 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSERANRSDLGAIYVCVG 180  
Db 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSERANRSDLGAIYVCVG 180  
Qy 181 KDFNETOLARIADSKDHVPVNDGFOALQGIHLSILKKSCEIILAAEPSTICAGESFOVV 240  
Db 181 KDFNETOLARIADSKDHVPVNDGFOALQGIHLSILKKSCEIILAAEPSTICAGESFOVV 240  
Qy 241 VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Qy 301 DGLSFSSSVIITTHCS 318  
Db 301 DGLSFSSSVIITTHCS 318  
RESULT 9  
PCT-US01-11988-621  
; Sequence 621, Application PC/TUS0111988  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: Albumin Fusion Proteins  
; FILE REFERENCE: PF546PCT  
; CURRENT APPLICATION NUMBER: PCT/US01/11988  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: 60/229, 358  
; PRIOR FILING DATE: 2000-04-12  
; PRIOR APPLICATION NUMBER: 60/256, 931  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/199, 384  
; PRIOR FILING DATE: 2000-04-25  
; NUMBER OF SEQ ID NOS: 2267  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 621  
; LENGTH: 403  
; TYPE: PRT

; ORGANISM: Homo sapiens  
PCT-US01-11988-621

Query Match 95.4%; Score 1649; DB 1; Length 403;  
Best Local Similarity 100.0%; Pred. No. 8.9e-160;  
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATAERRALGIGFOWLSLATLVLCAGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFOWLSLATLVLCAGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60  
Qy 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Db 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Qy 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSEREANRSDRLGAIVYCVGV 180  
Db 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSEREANRSDRLGAIVYCVGV 180  
Qy 181 KDFNETOLARIADSKDHVPVNDGFQALQGIHSHILKKSCEIILAAEPSTICAGESFQV 240  
Db 181 KDFNETOLARIADSKDHVPVNDGFQALQGIHSHILKKSCEIILAAEPSTICAGESFQV 240  
Qy 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGKMAALQVSMN 300  
Db 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGKMAALQVSMN 300  
Qy 301 DGLSFISSSVIITTHCS 318  
Db 301 DGLSFISSSVIITTHCS 318

RESULT 10

US-09-833-245-621  
; Sequence 621, Application US/09833245  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: Albumin Fusion Proteins  
; FILE REFERENCE: PF546PCT  
; CURRENT APPLICATION NUMBER: US/09/833,245  
; PRIOR FILING DATE: 2001-04-12  
; PRIOR APPLICATION NUMBER: 60/229, 358  
; PRIOR FILING DATE: 2000-04-12  
; PRIOR APPLICATION NUMBER: 60/256, 931  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/199, 384  
; PRIOR FILING DATE: 2000-04-25  
; NUMBER OF SEQ ID NOS: 2267  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 621  
; LENGTH: 403  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-833-245-621

Query Match 95.4%; Score 1649; DB 22; Length 403;  
Best Local Similarity 100.0%; Pred. No. 8.9e-160;  
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATAERRALGIGFOWLSLATLVLCAGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFOWLSLATLVLCAGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60  
Qy 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Db 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Qy 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSEREANRSDRLGAIVYCVGV 180  
Db 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSEREANRSDRLGAIVYCVGV 180

Qy 181 KDFNETOLARIADSKDHVPVNDGFQALQGIHSHILKKSCEIILAAEPSTICAGESFQV 240  
Db 181 KDFNETOLARIADSKDHVPVNDGFQALQGIHSHILKKSCEIILAAEPSTICAGESFQV 240  
Qy 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGKMAALQVSMN 300  
Db 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGKMAALQVSMN 300  
Qy 301 DGLSFISSSVIITTHCS 318  
Db 301 DGLSFISSSVIITTHCS 318

RESULT 11

US-09-918-715-187  
; Sequence 187, Application US/09918715  
; GENERAL INFORMATION:  
; APPLICANT: Brad St. Croix  
; APPLICANT: Bert Vogelstein  
; APPLICANT: Kenneth Kinzler  
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
; FILE REFERENCE: 1107.00134  
; CURRENT APPLICATION NUMBER: US/09/918,715  
; CURRENT FILING DATE: 2001-08-01  
; PRIOR APPLICATION NUMBER: 60/222,599  
; PRIOR FILING DATE: 2000-08-02  
; PRIOR APPLICATION NUMBER: 60/224,360  
; PRIOR FILING DATE: 2000-08-11  
; PRIOR APPLICATION NUMBER: 60/282,850  
; PRIOR FILING DATE: 2000-04-11  
; NUMBER OF SEQ ID NOS: 358  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 187  
; LENGTH: 564  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-918-715-187

Query Match 95.4%; Score 1649; DB 23; Length 564;  
Best Local Similarity 100.0%; Pred. No. 1.5e-159;  
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATAERRALGIGFOWLSLATLVLCAGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFOWLSLATLVLCAGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60  
Qy 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Db 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Qy 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSEREANRSDRLGAIVYCVGV 180  
Db 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSEREANRSDRLGAIVYCVGV 180  
Qy 181 KDFNETOLARIADSKDHVPVNDGFQALQGIHSHILKKSCEIILAAEPSTICAGESFQV 240  
Db 181 KDFNETOLARIADSKDHVPVNDGFQALQGIHSHILKKSCEIILAAEPSTICAGESFQV 240  
Qy 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGKMAALQVSMN 300  
Db 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGKMAALQVSMN 300  
Qy 301 DGLSFISSSVIITTHCS 318  
Db 301 DGLSFISSSVIITTHCS 318

RESULT 12

US-09-918-715-232  
; Sequence 232, Application US/09918715  
; GENERAL INFORMATION:  
; APPLICANT: Brad St. Croix

Fri Aug 9 10:57:07 2002

us-09-970-076-8.rapm

```

; APPLICANT: Bert Vogelstein
; APPLICANT: Kenneth Kinzler
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00134
; CURRENT APPLICATION NUMBER: US/09/918,715
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 60/222,599
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 60/224,360
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 232
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-918-715-232

Query Match          95.4%; Score 1649; DB 23; Length 564;
Best Local Similarity 100.0%; Pred. No. 1.5e-159;
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATAERRALGIGFOWLSLATLVLCAGOGGRRDGGPACYGCGFDLYFLDKSGSVLHHWN 60
Db 1 MATAERRALGIGFOWLSLATLVLCAGOGGRRDGGPACYGCGFDLYFLDKSGSVLHHWN 60
QY 61 EIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120
Db 61 EIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120
QY 121 HEGFERASEQIYYENRQGYRTASVIAITDGLHEDLFFYSERANRSDLGAIYVCVG 180
Db 121 HEGFERASEQIYYENRQGYRTASVIAITDGLHEDLFFYSERANRSDLGAIYVCVG 180
QY 181 KDFNETQLARIADSKDHVPVNDGFOALOGIIHSILKSKCIEILAAEPSTICAGESQV 240
Db 181 KDFNETQLARIADSKDHVPVNDGFOALOGIIHSILKSKCIEILAAEPSTICAGESQV 240
QY 241 VRGNGFRHARNVDVRLCSFKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300
Db 241 VRGNGFRHARNVDVRLCSFKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300
QY 301 DGLSFISSSVIITTHCS 318
Db 301 DGLSFISSSVIITTHCS 318

RESULT 14
PCT-US00-30045-125
; Sequence 125, Application PC/TUS0030045
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: PS708PCT
; CURRENT APPLICATION NUMBER: PCT/US00/30045
; CURRENT FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/163,581
; PRIOR FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: 60/215,133
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 201
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 125
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (175)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (320)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (331)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (368)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; PCT-US00-30045-125

Query Match          95.1%; Score 1644; DB 1; Length 403;
Best Local Similarity 99.7%; Pred. No. 2.9e-159;
Matches 317; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MATAERRALGIGFOWLSLATLVLCAGOGGRRDGGPACYGCGFDLYFLDKSGSVLHHWN 60
Db 1 MATAERRALGIGFOWLSLATLVLCAGOGGRRDGGPACYGCGFDLYFLDKSGSVLHHWN 60
QY 61 EIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120
```



Db 61 EIIYFVQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLELOKVLPGGDTYM 120  
Qy 121 HEGPERASEQIYYENROGYRTASVITIALTDGELHEDLFFYSEREANBSRDLGAIVYCVGV 180  
Db 121 HEGPERASEQIYYENROGYRTASVITIALTDGELHEDLFFYSEREANBSRDLGAIXYCVGV 180  
Qy 181 KDFNETOLARIADSKDHVPVNDGFQALQGIHLSILKKSCIEILAAEPSTICAGESFQV 240  
Db 181 KDFNETOLARIADSKDHVPVNDGFQALQGIHLSILKKSCIEILAAEPSTICAGESFQV 240  
Qy 241 VRNGFPHARNVDRVLCSEKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRNGFPHARNVDRVLCSEKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Qy 301 DGLSFSSSVIITTHCS 318  
Db 301 DGLSFSSSVIITTHCS 318

RESULT 15  
PCT-US01-11988-620  
; Sequence 620, Application PC/TUS0111988  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: Albumin Fusion Proteins  
; FILE REFERENCE: PF546PCT  
; CURRENT APPLICATION NUMBER: PCT/US01/11988  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: 60/229, 358  
; PRIOR FILING DATE: 2000-04-12  
; PRIOR APPLICATION NUMBER: 60/256, 931  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/199, 384  
; PRIOR FILING DATE: 2000-04-25  
; NUMBER OF SEQ ID NOS: 2267  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 620  
; LENGTH: 403  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (175)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (320)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (331)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (368)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
PCT-US01-11988-620

Query Match 95.18; Score 1644; DB 1; Length 403;  
Best Local Similarity 99.7%; Pred. No. 2.9e-159;  
Matches 317; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 MATERRALGIGFOWLSLATLVLCAGOGRRDGGPACYGFDLYFLDKSGSVLHHWN 60  
Db 1 MATERRALGIGFOWLSLATLVLCAGOGRRDGGPACYGFDLYFLDKSGSVLHHWN 60  
Qy 61 EIIYFVQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLELOKVLPGGDTYM 120  
Db 61 EIIYFVQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLELOKVLPGGDTYM 120  
Qy 121 HEGPERASEQIYYENROGYRTASVITIALTDGELHEDLFFYSEREANBSRDLGAIVYCVGV 180  
Db 121 HEGPERASEQIYYENROGYRTASVITIALTDGELHEDLFFYSEREANBSRDLGAIXYCVGV 180

Qy 181 KDFNETOLARIADSKDHVPVNDGFQALQGIHLSILKKSCIEILAAEPSTICAGESFQV 240  
Db 181 KDFNETOLARIADSKDHVPVNDGFQALQGIHLSILKKSCIEILAAEPSTICAGESFQV 240  
Qy 241 VRNGFPHARNVDRVLCSEKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRNGFPHARNVDRVLCSEKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Qy 301 DGLSFSSSVIITTHCS 318  
Db 301 DGLSFSSSVIITTHCS 318

Search completed: August 9, 2002, 10:44:32  
Job time: 870 sec

us-09-970-076-8.rapm

Fri Aug 9 10:57:07 2002

-----

```

Query Match      100.0%; Score 1914; DB 23; Length 368;
Best Local Similarity 100.0%; Pred. No. 3.le-179;
Matches 368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATERRALGIGFOWLSLATLVLCACGGGRREDGGPACYGDFLYFLDKSGSVLHHWN 60
      |||
Db: 1 MATERRALGIGFOWLSLATLVLCACGGGRREDGGPACYGDFLYFLDKSGSVLHHWN 60
      |||

```

Fri Aug 9 10:56:59 2002

```

Db 241 VRNGFRHARNVDRVLCSPKINDSVTLAEKPFSEDTYLLCPAPILKEVGKKAALQVSMN 300
QY 301 DGLSFSSSSVIITTHCSDGSLAIALLLFLLALALWFWPLCCCTVIKEVPPPPAE 360
Db 301 DGLSFSSSSVIITTHCSDGSLAIALLLFLLALALWFWPLCCCTVIKEVPPPPAE 360
QY 361 ESEK 364
Db 361 ESEE 364

```

RESULT 3

```

US-09-918-715-232
; Sequence 232, Application US/09918715
; GENERAL INFORMATION:
; APPLICANT: Brad St. Croix
; APPLICANT: Bert Vogelstein
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00134
; CURRENT APPLICATION NUMBER: 60/222,599
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 60/224,360
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 232
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-918-715-232

```

```

Query Match 99.0%; Score 1894; DB 23; Length 564;
Best Local Similarity 100.0%; Pred. No. 5.4e-177;
Matches 364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MATAERRALGIGFQWLATLVLCAGQGRREDGGPACYGFDLYFILDKSGSVLHHWN 60
Db 1 MATAERRALGIGFQWLATLVLCAGQGRREDGGPACYGFDLYFILDKSGSVLHHWN 60
QY 61 EYIFVEQLAHKFISPOLRMSFIVFSTRGTTMLKLTEDREQIROGLEELQKVLPGDITYM 120
Db 61 EYIFVEQLAHKFISPOLRMSFIVFSTRGTTMLKLTEDREQIROGLEELQKVLPGDITYM 120
QY 121 HEGFERASEQIYYENRQGYRTASVIALTDGELHEDLFFYSERANRSDGAIYVCVGV 180
Db 121 HEGFERASEQIYYENRQGYRTASVIALTDGELHEDLFFYSERANRSDGAIYVCVGV 180
QY 181 KDFNETQIARIADSKDHVPVNDGFOALQGIHSLKKSCEIILAAEPSTICAGESFQVY 240
Db 181 KDFNETQIARIADSKDHVPVNDGFOALQGIHSLKKSCEIILAAEPSTICAGESFQVY 240
QY 241 VRNGFRHARNVDRVLCSPKINDSVTLNEKPFSEDTYLLCPAPILKEVGKKAALQVSMN 300
Db 241 VRNGFRHARNVDRVLCSPKINDSVTLNEKPFSEDTYLLCPAPILKEVGKKAALQVSMN 300
QY 301 DGLSFSSSSVIITTHCSDGSLAIALLLFLLALALWFWPLCCCTVIKEVPPPPAE 360
Db 301 DGLSFSSSSVIITTHCSDGSLAIALLLFLLALALWFWPLCCCTVIKEVPPPPAE 360
QY 361 ESEK 364
Db 361 ESEE 364

```

RESULT 4  
US-09-970-076-6

```

QY 61 EYIFVEQLAHKFISPOLRMSFIVFSTRGTTMLKLTEDREQIROGLEELQKVLPGDITYM 120
Db 61 EYIFVEQLAHKFISPOLRMSFIVFSTRGTTMLKLTEDREQIROGLEELQKVLPGDITYM 120
QY 121 HEGFERASEQIYYENRQGYRTASVIALTDGELHEDLFFYSERANRSDGAIYVCVGV 180
Db 121 HEGFERASEQIYYENRQGYRTASVIALTDGELHEDLFFYSERANRSDGAIYVCVGV 180
QY 181 KDFNETQIARIADSKDHVPVNDGFOALQGIHSLKKSCEIILAAEPSTICAGESFQVY 240
Db 181 KDFNETQIARIADSKDHVPVNDGFOALQGIHSLKKSCEIILAAEPSTICAGESFQVY 240
QY 241 VRNGFRHARNVDRVLCSPKINDSVTLNEKPFSEDTYLLCPAPILKEVGKKAALQVSMN 300
Db 241 VRNGFRHARNVDRVLCSPKINDSVTLNEKPFSEDTYLLCPAPILKEVGKKAALQVSMN 300
QY 301 DGLSFSSSSVIITTHCSDGSLAIALLLFLLALALWFWPLCCCTVIKEVPPPPAE 360
Db 301 DGLSFSSSSVIITTHCSDGSLAIALLLFLLALALWFWPLCCCTVIKEVPPPPAE 360
QY 361 ESEK 368
Db 361 ESEK 368

```

RESULT 2

```

US-09-918-715-187
; Sequence 187, Application US/09918715
; GENERAL INFORMATION:
; APPLICANT: Brad St. Croix
; APPLICANT: Bert Vogelstein
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00134
; CURRENT APPLICATION NUMBER: 60/222,599
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 60/224,360
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 187
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-918-715-187

```

```

Query Match 99.0%; Score 1894; DB 23; Length 564;
Best Local Similarity 100.0%; Pred. No. 5.4e-177;
Matches 364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MATAERRALGIGFQWLATLVLCAGQGRREDGGPACYGFDLYFILDKSGSVLHHWN 60
Db 1 MATAERRALGIGFQWLATLVLCAGQGRREDGGPACYGFDLYFILDKSGSVLHHWN 60
QY 61 EYIFVEQLAHKFISPOLRMSFIVFSTRGTTMLKLTEDREQIROGLEELQKVLPGDITYM 120
Db 61 EYIFVEQLAHKFISPOLRMSFIVFSTRGTTMLKLTEDREQIROGLEELQKVLPGDITYM 120
QY 121 HEGFERASEQIYYENRQGYRTASVIALTDGELHEDLFFYSERANRSDGAIYVCVGV 180
Db 121 HEGFERASEQIYYENRQGYRTASVIALTDGELHEDLFFYSERANRSDGAIYVCVGV 180
QY 181 KDFNETQIARIADSKDHVPVNDGFOALQGIHSLKKSCEIILAAEPSTICAGESFQVY 240
Db 181 KDFNETQIARIADSKDHVPVNDGFOALQGIHSLKKSCEIILAAEPSTICAGESFQVY 240
QY 241 VRNGFRHARNVDRVLCSPKINDSVTLNEKPFSEDTYLLCPAPILKEVGKKAALQVSMN 300

```

; Sequence 6, Application US/09970076  
; GENERAL INFORMATION:  
; APPLICANT: Young, John A.T.  
; APPLICANT: Bradley, Kenneth A.  
; APPLICANT: Collier, Robert J.  
; APPLICANT: Mogridge, Jeremy S.  
; TITLE OF INVENTION: Anthrax Toxin Receptor  
; FILE REFERENCE: 960296.97745  
; CURRENT APPLICATION NUMBER: US/09/970,076  
; CURRENT FILING DATE: 2001-10-03  
; PRIOR APPLICATION NUMBER: 60/251,481  
; PRIOR FILING DATE: 2000-12-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 564  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-970-076-6

Query Match 99.0%; Score 1894; DB 23; Length 564;  
Best Local Similarity 100.0%; Pred. No. 5.4e-177;  
Matches 364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MATAERRALGIGFOWLSLATLVLCAGOGGRRDGGPGACYGDFLYFLDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFOWLSLATLVLCAGOGGRRDGGPGACYGDFLYFLDKSGSVLHHWN 60  
Qy 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120  
Db 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120  
Qy 121 HEGFERASEQIYYENRQGYRTASVIIALTGDGELHEDLFFYSERANRSDLGAIYVCV 180  
Db 121 HEGFERASEQIYYENRQGYRTASVIIALTGDGELHEDLFFYSERANRSDLGAIYVCV 180  
Qy 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHSLTKKSCIEILAAEPSTICAGESFOV 240  
Db 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHSLTKKSCIEILAAEPSTICAGESFOV 240  
Qy 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFVSVDYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFVSVDYLLCPAPILKEVGMKAALQVSMN 300  
Qy 301 DGLSFISSVIITTHCSGDSILATALLILFLLALALLMFWPLCCCTVIIKEVPPPPAE 360  
Db 301 DGLSFISSVIITTHCSGDSILATALLILFLLALALLMFWPLCCCTVIIKEVPPPPAE 360  
Qy 361 ESEE 364  
Db 361 ESEE 364

RESULT 5  
PCT-US00-30045-94  
; Sequence 94, Application PC/TUS0030045  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: 28 Human Secreted Proteins  
; FILE REFERENCE: PS708PCT  
; CURRENT APPLICATION NUMBER: PCT/US00/30045  
; CURRENT FILING DATE: 2000-11-01  
; PRIOR APPLICATION NUMBER: 60/163,581  
; PRIOR FILING DATE: 1999-11-05  
; PRIOR APPLICATION NUMBER: 60/215,133  
; PRIOR FILING DATE: 2000-06-30  
; NUMBER OF SEQ ID NOS: 201  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 94  
; LENGTH: 403  
; TYPE: PRT  
; ORGANISM: Homo sapiens

PCT-US00-30045-94

Query Match 98.7%; Score 1889; DB 1; Length 403;  
Best Local Similarity 100.0%; Pred. No. 1e-176;  
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MATAERRALGIGFOWLSLATLVLCAGOGGRRDGGPGACYGDFLYFLDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFOWLSLATLVLCAGOGGRRDGGPGACYGDFLYFLDKSGSVLHHWN 60  
Qy 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120  
Db 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120  
Qy 121 HEGFERASEQIYYENRQGYRTASVIIALTGDGELHEDLFFYSERANRSDLGAIYVCV 180  
Db 121 HEGFERASEQIYYENRQGYRTASVIIALTGDGELHEDLFFYSERANRSDLGAIYVCV 180  
Qy 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHSLTKKSCIEILAAEPSTICAGESFOV 240  
Db 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHSLTKKSCIEILAAEPSTICAGESFOV 240  
Qy 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFVSVDYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFVSVDYLLCPAPILKEVGMKAALQVSMN 300  
Qy 301 DGLSFISSVIITTHCSGDSILATALLILFLLALALLMFWPLCCCTVIIKEVPPPPAE 360  
Db 301 DGLSFISSVIITTHCSGDSILATALLILFLLALALLMFWPLCCCTVIIKEVPPPPAE 360  
Qy 361 ESE 363  
Db 361 ESE 363

RESULT 6  
PCT-US01-11988-621  
; Sequence 621, Application PC/TUS0111988  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: Albumin Fusion Proteins  
; FILE REFERENCE: PF546PCT  
; CURRENT APPLICATION NUMBER: PCT/US01/11988  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: 60/229, 358  
; PRIOR FILING DATE: 2000-04-12  
; PRIOR APPLICATION NUMBER: 60/256, 931  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/199, 384  
; PRIOR FILING DATE: 2000-04-25  
; NUMBER OF SEQ ID NOS: 2267  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 621  
; LENGTH: 403  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US01-11988-621

Query Match 98.7%; Score 1889; DB 1; Length 403;  
Best Local Similarity 100.0%; Pred. No. 1e-176;  
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MATAERRALGIGFOWLSLATLVLCAGOGGRRDGGPGACYGDFLYFLDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFOWLSLATLVLCAGOGGRRDGGPGACYGDFLYFLDKSGSVLHHWN 60  
Qy 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120  
Db 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120  
Qy 121 HEGFERASEQIYYENRQGYRTASVIIALTGDGELHEDLFFYSERANRSDLGAIYVCV 180

Fri Aug 9 10:56:59 2002

```
|||||
Db 121 HEGFERASEQIYYENRQGYRTASVIALTDGELHEDLFFYSEREANRSRDLGAIVYCVG 180
QY 181 KDFNETOLARIADSKDHVFPVNDGFQALOGIIHSILKKSCIEIIAAEPSTICAGESFQV 240
Db 181 KDFNETOLARIADSKDHVFPVNDGFQALOGIIHSILKKSCIEIIAAEPSTICAGESFQV 240
QY 241 VRNGFRHARNVDRVLCSPKINDSVTLNEKPFVSEDYTLVLLCPAPILKEVGMKAALQVSMN 300
Db 241 VRNGFRHARNVDRVLCSPKINDSVTLNEKPFVSEDYTLVLLCPAPILKEVGMKAALQVSMN 300
QY 301 DGLSFSSSVIITTHCSDGSILAIALLILFLALLALWFWPLCCTVIIVKEVPPPPAE 360
Db 301 DGLSFSSSVIITTHCSDGSILAIALLILFLALLALWFWPLCCTVIIVKEVPPPPAE 360
QY 361 ESE 363
Db 361 ESE 363

RESULT 7
US-09-833-245-621
; Sequence 621, Application US/09833245
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229, 358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256, 931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199, 384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 621
; TYPE: PRT
; LENGTH: 403
; ORGANISM: Homo sapiens
US-09-833-245-621

Query Match 98.7%; Score 1889; DB 22; Length 403;
Best Local Similarity 100.0%; Pred. No. 1e-176;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATERRALGIGFOWLSLATLVLCAGGGRRDGGPACYGDFLYFILDKSGSVLHHN 60
Db 1 MATERRALGIGFOWLSLATLVLCAGGGRRDGGPACYGDFLYFILDKSGSVLHHN 60
QY 61 EIYFVEQLAHKFTSPQLRMSFVIFSTRGTTLMKLTEDREQIRQGLELQKVLPGGDTYM 120
Db 61 EIYFVEQLAHKFTSPQLRMSFVIFSTRGTTLMKLTEDREQIRQGLELQKVLPGGDTYM 120
QY 121 HEGFERASEQIYYENRQGYRTASVIALTDGELHEDLFFYSEREANRSRDLGAIVYCVG 180
Db 121 HEGFERASEQIYYENRQGYRTASVIALTDGELHEDLFFYSEREANRSRDLGAIVYCVG 180
QY 181 KDFNETOLARIADSKDHVFPVNDGFQALOGIIHSILKKSCIEIIAAEPSTICAGESFQV 240
Db 181 KDFNETOLARIADSKDHVFPVNDGFQALOGIIHSILKKSCIEIIAAEPSTICAGESFQV 240
QY 241 VRNGFRHARNVDRVLCSPKINDSVTLNEKPFVSEDYTLVLLCPAPILKEVGMKAALQVSMN 300
Db 241 VRNGFRHARNVDRVLCSPKINDSVTLNEKPFVSEDYTLVLLCPAPILKEVGMKAALQVSMN 300
QY 301 DGLSFSSSVIITTHCSDGSILAIALLILFLALLALWFWPLCCTVIIVKEVPPPPAE 360
Db 301 DGLSFSSSVIITTHCSDGSILAIALLILFLALLALWFWPLCCTVIIVKEVPPPPAE 360
QY 361 ESE 363
```

```
|||||
Db 361 ESE 363

RESULT 8
PCT-US00-30045-125
; Sequence 125, Application PC/TUS0030045
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: PS708PCT
; CURRENT APPLICATION NUMBER: PCT/US00/30045
; CURRENT FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/163,581
; PRIOR FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: 60/215,133
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 201
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 125
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (175)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (320)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (331)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (368)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
PCT-US00-30045-125

Query Match 97.7%; Score 1870; DB 1; Length 403;
Best Local Similarity 99.2%; Pred. No. 7.7e-175;
Matches 360; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MATERRALGIGFOWLSLATLVLCAGGGRRDGGPACYGDFLYFILDKSGSVLHHN 60
Db 1 MATERRALGIGFOWLSLATLVLCAGGGRRDGGPACYGDFLYFILDKSGSVLHHN 60
QY 61 EIYFVEQLAHKFTSPQLRMSFVIFSTRGTTLMKLTEDREQIRQGLELQKVLPGGDTYM 120
Db 61 EIYFVEQLAHKFTSPQLRMSFVIFSTRGTTLMKLTEDREQIRQGLELQKVLPGGDTYM 120
QY 121 HEGFERASEQIYYENRQGYRTASVIALTDGELHEDLFFYSEREANRSRDLGAIVYCVG 180
Db 121 HEGFERASEQIYYENRQGYRTASVIALTDGELHEDLFFYSEREANRSRDLGAIVYCVG 180
QY 181 KDFNETOLARIADSKDHVFPVNDGFQALOGIIHSILKKSCIEIIAAEPSTICAGESFQV 240
Db 181 KDFNETOLARIADSKDHVFPVNDGFQALOGIIHSILKKSCIEIIAAEPSTICAGESFQV 240
QY 241 VRNGFRHARNVDRVLCSPKINDSVTLNEKPFVSEDYTLVLLCPAPILKEVGMKAALQVSMN 300
Db 241 VRNGFRHARNVDRVLCSPKINDSVTLNEKPFVSEDYTLVLLCPAPILKEVGMKAALQVSMN 300
QY 301 DGLSFSSSVIITTHCSDGSILAIALLILFLALLALWFWPLCCTVIIVKEVPPPPAE 360
Db 301 DGLSFSSSVIITTHCSDGSILAIALLILFLALLALWFWPLCCTVIIVKEVPPPPAE 360
QY 361 ESE 363
Db 361 ESE 363

RESULT 9
```



[illegible]



Db 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Qy 301 DGLSFSSSVIITTHCS 318  
Db 301 DGLSFSSSVIITTHCS 318  
RESULT 14  
US-09-223-546-9  
; Sequence 9, Application US/09223546  
; GENERAL INFORMATION:  
; APPLICANT: Holtzman, Douglas  
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM  
; FILE REFERENCE: 09404/066001  
; CURRENT APPLICATION NUMBER: US/09/223,546  
; CURRENT FILING DATE: 1998-12-30  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 9  
; LENGTH: 333  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-223-546-9

Query Match 86.2%; Score 1649; DB 16; Length 333;  
Best Local Similarity 100.0%; Pred. No. 3.5e-153;  
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATAERRALGIGFQWLSLATLVLCACGGGRREDGGPACYGFDLYFTLDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFQWLSLATLVLCACGGGRREDGGPACYGFDLYFTLDKSGSVLHHWN 60  
Qy 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMLKLTEDREQIROGLELQKVLPGGDTYM 120  
Db 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMLKLTEDREQIROGLELQKVLPGGDTYM 120  
Qy 121 HEGFERASEQIYYENROGYRTASVITIALTDGELHEDLFFYSEREANRSRDLGAIVYCVGV 180  
Db 121 HEGFERASEQIYYENROGYRTASVITIALTDGELHEDLFFYSEREANRSRDLGAIVYCVGV 180  
Qy 181 KDFNETQIARIADSKDHVPVNDGFQALQGIHSLKKSCIEILAAEPSTICAGESFOVV 240  
Db 181 KDFNETQIARIADSKDHVPVNDGFQALQGIHSLKKSCIEILAAEPSTICAGESFOVV 240  
Qy 241 VRNGCFRHRARNVDRVLCFSKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRNGCFRHRARNVDRVLCFSKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Qy 301 DGLSFSSSVIITTHCS 318  
Db 301 DGLSFSSSVIITTHCS 318

RESULT 15  
US-09-471-179-9  
; Sequence 9, Application US/09471179  
; GENERAL INFORMATION:  
; APPLICANT: Holtzman, Douglas  
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM  
; FILE REFERENCE: 7853-173  
; CURRENT APPLICATION NUMBER: US/09/471,179  
; CURRENT FILING DATE: 1999-12-23  
; PRIOR APPLICATION NUMBER: 09/223,546  
; PRIOR FILING DATE: 1998-12-30  
; NUMBER OF SEQ ID NOS: 135  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 333  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-471-179-9

Query Match 86.2%; Score 1649; DB 18; Length 333;  
Best Local Similarity 100.0%; Pred. No. 3.5e-153;  
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MATAERRALGIGFQWLSLATLVLCACGGGRREDGGPACYGFDLYFTLDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFQWLSLATLVLCACGGGRREDGGPACYGFDLYFTLDKSGSVLHHWN 60  
Qy 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMLKLTEDREQIROGLELQKVLPGGDTYM 120  
Db 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMLKLTEDREQIROGLELQKVLPGGDTYM 120  
Qy 121 HEGFERASEQIYYENROGYRTASVITIALTDGELHEDLFFYSEREANRSRDLGAIVYCVGV 180  
Db 121 HEGFERASEQIYYENROGYRTASVITIALTDGELHEDLFFYSEREANRSRDLGAIVYCVGV 180  
Qy 181 KDFNETQIARIADSKDHVPVNDGFQALQGIHSLKKSCIEILAAEPSTICAGESFOVV 240  
Db 181 KDFNETQIARIADSKDHVPVNDGFQALQGIHSLKKSCIEILAAEPSTICAGESFOVV 240  
Qy 241 VRNGCFRHRARNVDRVLCFSKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRNGCFRHRARNVDRVLCFSKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Qy 301 DGLSFSSSVIITTHCS 318  
Db 301 DGLSFSSSVIITTHCS 318

Search completed: August 9, 2002, 10:44:29  
Job time: 867 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:33:10 ; Search time 45.48 Seconds  
(without alignments)  
117.080 Million cell updates/sec

Title: US-09-970-076-10

Perfect score: 1131

Sequence: 1 MSFVFSFTRCTTLMLKLTDR.....STSGFKEGNSHPCLPARPHT 218

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCUTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	128.5	11.4	1155	1 US-08-286-889-46	Sequence 46, Appl
2	128.5	11.4	1155	1 US-08-485-618-46	Sequence 46, Appl
3	128.5	11.4	1155	1 US-08-362-652-46	Sequence 46, Appl
4	128.5	11.4	1155	2 US-08-605-672-46	Sequence 46, Appl
5	128.5	11.4	1155	2 US-08-482-293A-46	Sequence 46, Appl
6	128.5	11.4	1155	2 US-08-943-363-46	Sequence 46, Appl
7	128.5	11.4	1155	4 US-09-193-043-46	Sequence 46, Appl
8	128.5	11.4	1161	1 US-08-485-618-53	Sequence 53, Appl
9	128.5	11.4	1161	1 US-08-362-652-53	Sequence 53, Appl
10	128.5	11.4	1161	2 US-08-605-672-53	Sequence 53, Appl
11	128.5	11.4	1161	2 US-08-482-293A-53	Sequence 53, Appl
12	128.5	11.4	1161	2 US-08-943-363-53	Sequence 53, Appl
13	128.5	11.4	1161	4 US-09-193-043-53	Sequence 53, Appl
14	122	10.8	1151	1 US-08-286-889-37	Sequence 37, Appl
15	122	10.8	1151	1 US-08-485-618-37	Sequence 37, Appl
16	122	10.8	1151	1 US-08-362-652-37	Sequence 37, Appl
17	122	10.8	1151	2 US-08-605-672-37	Sequence 37, Appl
18	122	10.8	1151	2 US-08-482-293A-37	Sequence 37, Appl
19	122	10.8	1151	4 US-09-193-043-37	Sequence 37, Appl
20	122	10.8	1151	4 US-09-193-043-37	Sequence 37, Appl
21	122	10.8	1161	1 US-08-485-618-55	Sequence 55, Appl
22	122	10.8	1161	1 US-08-362-652-55	Sequence 55, Appl
23	122	10.8	1161	2 US-08-605-672-55	Sequence 55, Appl
24	122	10.8	1161	2 US-08-482-293A-55	Sequence 55, Appl
25	122	10.8	1161	2 US-08-943-363-55	Sequence 55, Appl
26	122	10.8	1161	4 US-09-193-043-55	Sequence 55, Appl
27	114	10.1	1161	1 US-08-173-497-2	Sequence 2, Appl1

28 114 10.1 1161 1 US-08-286-889-2 Sequence 2, Appl1  
29 114 10.1 1161 1 US-08-485-618-2 Sequence 2, Appl1  
30 114 10.1 1161 1 US-08-485-618-99 Sequence 99, Appl  
31 114 10.1 1161 1 US-08-362-652-2 Sequence 2, Appl1  
32 114 10.1 1161 2 US-08-605-672-2 Sequence 99, Appl  
33 114 10.1 1161 2 US-08-605-672-99 Sequence 2, Appl1  
34 114 10.1 1161 2 US-08-482-293A-2 Sequence 99, Appl  
35 114 10.1 1161 2 US-08-482-293A-99 Sequence 2, Appl1  
36 114 10.1 1161 2 US-08-943-363-2 Sequence 99, Appl  
37 114 10.1 1161 2 US-08-943-363-99 Sequence 2, Appl1  
38 114 10.1 1161 4 US-09-193-043-2 Sequence 99, Appl  
39 114 10.1 1161 4 US-09-193-043-99 Sequence 43, Appl  
40 111 9.8 1152 2 US-08-476-062A-43 Sequence 43, Appl  
41 111 9.8 1152 5 PCT-US96-01314-43 Patent No. 5424399  
42 111 9.8 1152 6 5424399-2 Sequence 3, Appl1  
43 111 9.8 1153 1 US-08-173-497-3 Sequence 3, Appl1  
44 111 9.8 1153 1 US-08-286-889-3 Sequence 3, Appl1  
45 111 9.8 1153 1 US-08-485-618-3 Sequence 3, Appl1

#### ALIGNMENTS

RESULT 1

US-08-286-889-46

; Sequence 46, Application US/08286889

; Patent No. 5470953

; GENERAL INFORMATION:

; APPLICANT: Gallatin, W. Mich

; APPLICANT: Van der Vieren, Monica

; TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit

; NUMBER OF SEQUENCES: 51

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

; STREET: 233 South Wacker Drive, 6300 Sear Tower

; CITY: Chicago

; STATE: Illinois

; COUNTRY: United States

; ZIP: 60606-6402

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/286,889

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/173,497

; FILING DATE: 23-DEC-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Williams Jr., Joseph A.

; REGISTRATION NUMBER: P38,659

; REFERENCE/DOCKET NUMBER: 27866/32168

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312-474-6300

; TELEFAX: 312-474-0448

; TELEX: 25-3856

; INFORMATION FOR SEQ ID NO: 46:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1155 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-286-889-46

Query Match 11.4%; Score 128.5; DB 1; Length 1155;

Best Local Similarity 28.9%; Pred. No. 6.8e-06;  
Matches 50; Conservative 31; Mismatches 69; Indels 23; Gaps 9;

QY 12 TLMKLTEDRIQRLGLEELQKVLPGGDTYMHEGFRASEQIYYENRQGYRTA-SVIALT 70

Fri Aug 9 10:56:54 2002

```

Db 205 TEFKSSLSPOSILVDAIVQLQ-----GLTYTASGIQKVVVKELFHSKNGARKSAKKILIVIT 259
QY 71 DGEHEDLFYSE--REANRSDIGAIVYGVKVD-FNE-TOLARI-----ADSKDHVFP 121
Db 260 DQKFRDPLEYRHVPIPEAKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSQDHVFK 316
QY 122 VNDGFOALQGIHSLTKKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVD 174
Db 317 VGN-FVALKRSIQRIQEK-----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

RESULT 2
US-08-485-618-46
; Sequence 46, Application US/08485618
; Patent No. 5728533
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,618
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32797
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1155 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-485-618-46

Query Match 11.4%; Score 128.5; DB 1; Length 1155;
Best Local Similarity 28.9%; Pred. No. 6.8e-06;
Matches 50; Conservative 31; Mismatches 69; Indels 23; Gaps 9;

QY 12 TLMLKLTREDRIQGLEELQKVLPGGDTYMHGFEFASQIYYENRQGYRTA-SVIALT 70
Db 205 TEFKSSLSPOSILVDAIVQLQ-----GLTYTASGIQKVVVKELFHSKNGARKSAKKILIVIT 259
QY 71 DGEHEDLFYSE--REANRSDIGAIVYGVKVD-FNE-TOLARI-----ADSKDHVFP 121
Db 260 DQKFRDPLEYRHVPIPEAKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSQDHVFK 316
QY 122 VNDGFOALQGIHSLTKKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVD 174
Db 317 VGN-FVALKRSIQRIQEK-----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

RESULT 4

```

```

Db 260 DQKFRDPLEYRHVPIPEAKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSQDHVFK 316
QY 122 VNDGFOALQGIHSLTKKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVD 174
Db 317 VGN-FVALKRSIQRIQEK-----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

RESULT 3
US-08-362-652-46
; Sequence 46, Application US/08362652
; Patent No. 5766850
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,652
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32391
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1155 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-362-652-46

Query Match 11.4%; Score 128.5; DB 1; Length 1155;
Best Local Similarity 28.9%; Pred. No. 6.8e-06;
Matches 50; Conservative 31; Mismatches 69; Indels 23; Gaps 9;

QY 12 TLMLKLTREDRIQGLEELQKVLPGGDTYMHGFEFASQIYYENRQGYRTA-SVIALT 70
Db 205 TEFKSSLSPOSILVDAIVQLQ-----GLTYTASGIQKVVVKELFHSKNGARKSAKKILIVIT 259
QY 71 DGEHEDLFYSE--REANRSDIGAIVYGVKVD-FNE-TOLARI-----ADSKDHVFP 121
Db 260 DQKFRDPLEYRHVPIPEAKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSQDHVFK 316
QY 122 VNDGFOALQGIHSLTKKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVD 174
Db 317 VGN-FVALKRSIQRIQEK-----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

RESULT 4

```



```

STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,363
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1155 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-943-363-46

Query Match 11.4%; Score 128.5; DB 2; Length 1155;
Best Local Similarity 28.9%; Pred. No. 6.8e-06;
Matches 50; Conservative 31; Mismatches 69; Indels 23; Gaps 9;

QY 12 TLMKLTEDREQTQGLGELQKVLPGGDTYMHEGPERASEQIYYENRQGVRTA-SVIALT 70
| | : : : : : | | : : : : : | : : : : : | : : : : : | : : : : :
Db 205 TEFKSLSPQSLVDIAVLQ-----GLTYTASGQKVQVKELFHSKNGARKSAKKLIVIT 259
| | : : : : : | | : : : : : | : : : : : | : : : : : | : : : : :
QY 71 DGLHEDLPFYSE--REANRSRDGAIVYGVVD-FNE-TQLARI-----ADSKDHVP 121
| | : : : : : | | : : : : : | : : : : : | : : : : : | : : : : :
Db 260 DGQKRPDPLEYRHVTPAEKA---GIIPYAIGVGDAFREPTAQELWTICASQDHVPK 316
| | : : : : : | | : : : : : | : : : : : | : : : : : | : : : : :
QY 122 VNDGFOALOGIIHSLKSCIEILAAEPTICAGESFOFVVRGNGFHRARNV 174
| | : : : : : | | : : : : : | : : : : : | : : : : : | : : : : :
Db 317 VGN-FVALRSIQRIQEK-----IFAIEGTESRSSSSFOHEMSQGFSSALSMD 364

RESULT 7
US-09-193-043-46
; Sequence 46, Application US/09193043
; Patent No. 6251395
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 6251395el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/193,043
; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173,497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286,889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362,652

```

LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-485-618-53

Query Match 11.4%; Score 128.5; DB 1; Length 1161;  
Best Local Similarity 28.9%; Pred. No. 6.9e-06;  
Matches 50; Conservative 31; Mismatches 69; Indels 23; Gaps 9;  
QY 12 TLMKLTEDREQIRQGLEELQKVLPGGDTYMHGFEFASQIYYENRQGYRTA-SVIAALT 70  
DB 205 TEFKSSLSQSLVDAIVQLQ-----GLTYTASGIQVVKELFHSKNGARKSAKKILIVIT 259  
QY 71 DGEHLEDLFYSE--REANRSRDGLAIVYCVGKVD-FNE-TOLARI-----ADSKDHVFP 121  
DB 260 DGQKFRDPLEYRHVPEAKA---GIIRYAIGVDADFREPTALQELNTIGSAPSQDHVFK 316  
QY 122 VNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQVVRNGRFRHARNVD 174  
DB 317 VGN-FVALRSIQRIQIEK----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

RESULT 9  
US-08-362-652-53  
; Sequence 53, Application US/08362652  
; Patent No. 5766850  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/362,652  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32391  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1161 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-362-652-53

Query Match 11.4%; Score 128.5; DB 1; Length 1161;

Best Local Similarity 28.9%; Pred. No. 6.9e-06;  
Matches 50; Conservative 31; Mismatches 69; Indels 23; Gaps 9;  
QY 12 TLMKLTEDREQIRQGLEELQKVLPGGDTYMHGFEFASQIYYENRQGYRTA-SVIAALT 70  
DB 205 TEFKSSLSQSLVDAIVQLQ-----GLTYTASGIQVVKELFHSKNGARKSAKKILIVIT 259  
QY 71 DGEHLEDLFYSE--REANRSRDGLAIVYCVGKVD-FNE-TOLARI-----ADSKDHVFP 121  
DB 260 DGQKFRDPLEYRHVPEAKA---GIIRYAIGVDADFREPTALQELNTIGSAPSQDHVFK 316  
QY 122 VNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQVVRNGRFRHARNVD 174  
DB 317 VGN-FVALRSIQRIQIEK----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364  
RESULT 10  
US-08-605-672-53  
; Sequence 53, Application US/08605672  
; Patent No. 5817515  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/605,672  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32684  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1161 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-605-672-53

Query Match 11.4%; Score 128.5; DB 2; Length 1161;  
Best Local Similarity 28.9%; Pred. No. 6.9e-06;  
Matches 50; Conservative 31; Mismatches 69; Indels 23; Gaps 9;  
QY 12 TLMKLTEDREQIRQGLEELQKVLPGGDTYMHGFEFASQIYYENRQGYRTA-SVIAALT 70

us-09-970-076-10.ra1

Fri Aug 9 10:56:54 2002

Db 205 TEFKSSLSQSLVDAIVOLQ-----GLTYTASGIQKVVVKGKNGARKSAKILIVIT 259  
QY 71 DGEHEDLFFYSE--REANRSDLGAIYVGVKD-FNE-TQLARI-----ADSKDHVFP 121  
Db 260 DQKFRDPLEYRHHVPEAKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSDHVK 316  
QY 122 VNDGFOALOGIIHSLKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVD 174  
Db 317 VGN-FVALRSIQROIQEK-----IFAIEGTSRSSSFOHEMSQEGFSSALSMD 364

RESULT 11  
US-08-482-293A-53  
; Sequence 53, Application US/08482293A  
; Patent No. 5831029  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/482,293A  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32684  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1161 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-482-293A-53

Query Match 11.4%; Score 128.5; DB 2; Length 1161;  
Best Local Similarity 28.9%; Pred. No. 6.9e-06;  
Matches 50; Conservative 31; Mismatches 69; Indels 23; Gaps 9;  
QY 12 TLMLKLTREDRIQRLGLEELQKVLPGDGYMHGFEFASQIYYENRQGYRTA-SVILALT 70  
Db 205 TEFKSSLSQSLVDAIVOLQ-----GLTYTASGIQKVVVKGKNGARKSAKILIVIT 259  
QY 71 DGEHEDLFFYSE--REANRSDLGAIYVGVKD-FNE-TQLARI-----ADSKDHVFP 121  
Db 260 DQKFRDPLEYRHHVPEAKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSDHVK 316

QY 122 VNDGFOALOGIIHSLKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVD 174  
Db 317 VGN-FVALRSIQROIQEK-----IFAIEGTSRSSSFOHEMSQEGFSSALSMD 364  
RESULT 12  
US-08-943-363-53  
; Sequence 53, Application US/08943363  
; Patent No. 5837478  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/943,363  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32684  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1161 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-943-363-53

Query Match 11.4%; Score 128.5; DB 2; Length 1161;  
Best Local Similarity 28.9%; Pred. No. 6.9e-06;  
Matches 50; Conservative 31; Mismatches 69; Indels 23; Gaps 9;  
QY 12 TLMLKLTREDRIQRLGLEELQKVLPGDGYMHGFEFASQIYYENRQGYRTA-SVILALT 70  
Db 205 TEFKSSLSQSLVDAIVOLQ-----GLTYTASGIQKVVVKGKNGARKSAKILIVIT 259  
QY 71 DGEHEDLFFYSE--REANRSDLGAIYVGVKD-FNE-TQLARI-----ADSKDHVFP 121  
Db 260 DQKFRDPLEYRHHVPEAKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSDHVK 316  
QY 122 VNDGFOALOGIIHSLKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVD 174  
Db 317 VGN-FVALRSIQROIQEK-----IFAIEGTSRSSSFOHEMSQEGFSSALSMD 364



```

; FILING DATE: 23-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: P38,659
; REFERENCE/DOCKET NUMBER: 27866/32168
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-286-889-37

Query Match 10.8%; Score 122; DB 1; Length 1151;
Best Local Similarity 27.7%; Pred. No. 3.9e-05;
Matches 52; Conservative 32; Mismatches 76; Indels 28; Gaps

QY 2 SFIVFSTRGTTLMKLTET---DREQIRQGLEELQKVLPGGDTVMHEGFFERASEQIVYEN 56
DB 180 SLMOYSNLKTHFTTFEKNILDPQSLVDPIVQLQ-----GLTYTATGIRTVMEELFHSK 234
QY 57 RQGYRTA-SVIALTDELGHEDLFFYSE--REANRSRDLAGIYVCVGKVD-FNE-TQLAR 111
DB 235 NGSRSKAKKILLVITDQKYRDPLEYSDVIPADKA---GIIRYAIGVGDAFQEPPTALKE 291
QY 112 I-----ADSKHVPFVNDGFOALQGIITHSLKSCIBILAAEPSTICAGSFQVVVRGNG 166
DB 292 LNTIGSAPPQDHVFVKVGN-FAALRSIQRLQLEK----IFAIEGTQSRSSSFQHEMSQEG 346
QY 167 FRHARNVD 174
DB 347 FSSALTS D 354

RESULT 15
US-08-485-618-37
; Sequence 37, Application US/08485618
; Patent No. 5728533
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESSES:
; ADDRESSSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,618
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994

```

us-09-970-076-10.rai

Fri Aug 9 10:56:54 2002

```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 381659
; REFERENCE/DOCKET NUMBER: 27866/32797
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
;
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-485-618-37

Query Match      10.8%; Score 122; DB 1; Length 1151;
Best Local Similarity 27.7%; Pred. No. 3.9e-05;
Matches 52; Conservative 32; Mismatches 76; Indels 28; Gaps 10;

QY 2 SFIVFSTRGTTLMKLT-----DREIQRLGLEELQKVLPGDYMHEGFERASEQIYYEN 56
Db 180 SLMOQYSNILKTHFTTEFKNILDQSLVDPIVQLQ-----GLTYTATGIRTVMEELFHSK 234

QY 57 ROGYRTA-SVITATDGLHEDLFYSE--REANRSRDLGAIVYCVGVKD-FNE-TOLAR 111
Db 235 NGSRSKAKILLIVTDGOKYRDPLEYSDVIPADKA---GIIRYAIGVGDAFQEPALKE 291

QY 112 I-----ADSKDHVPVNDGFOALQGIHSILKSCIEILAAEPSTICAGESFQVVVRNG 166
Db 292 LNTIGSAPPQDHVFKVGN-FALRSIQRLQLEK---IFAIEGTQSRSSSSSQHEMSQEG 346

QY 167 FRHARNVD 174
Db 347 FSSALTS D 354
```

Search completed: August 9, 2002, 10:33:12  
Job time: 191 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:46:26 ; Search time 102.68 Seconds  
(without alignments)  
367.286 Million cell updates/sec

Title: US-09-970-076-10  
Perfect score: 1131  
Sequence: 1 MSFVFSFGTTLMLKLTDR.....STSGFKGNSHPCLPAPRHT 218

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	966	85.4	368	4 Q96P02	Q96P02 homo sapien
2	506	44.7	245	4 Q96NC7	Q96nc7 homo sapien
3	151	13.4	97	4 Q96EC6	Q96ec6 homo sapien
4	123	10.9	1332	5 Q9BPQ8	Q9bpq8 halocynthia
5	122	10.8	1161	11 Q9QVE7	Q9qve7 rattus norv
6	109	9.6	920	6 Q28984	Q28984 sus scrofa
7	107.5	9.5	517	4 Q43853	Q43853 homo sapien
8	106.5	9.4	760	11 Q70350	Q70350 mus musculus
9	105.5	9.3	724	5 Q04588	Q04588 elmeria max
10	104	9.2	205	11 Q63001	Q63001 rattus norv
11	103	9.1	660	4 Q9UGC3	Q9ugc3 homo sapien
12	103	9.1	1151	11 Q9JI30	Q9ji30 rattus norv
13	99	8.8	698	16 Q97IT1	Q97it1 cistridium
14	96.5	8.5	817	3 Q9US03	Q9us03 schizosacch
15	95	8.4	764	4 Q96HX6	Q96hx6 homo sapien
16	94.5	8.4	2944	11 Q63870	Q63870 mus musculus

17	92.5	8.2	1472	13 Q90ZA0	Q90za0 gallus gall
18	92.5	8.2	4307	5 Q19319	Q19319 caenorhabdi
19	92	8.1	937	4 Q96FT5	Q96ft5 homo sapien
20	92	8.1	1140	4 Q9P218	Q9p218 homo sapien
21	92	8.1	1207	4 Q9BQU7	Q9bqu7 homo sapien
22	91	8.0	675	10 Q9SG86	Q9sg86 arabidopsis
23	90.5	8.0	650	11 Q9CY21	Q9cy21 mus musculu
24	90	8.0	712	5 Q43981	Q43981 elmeria ten
25	89.5	7.9	1394	2 Q9AIU3	Q9aiu3 ehrlichia p
26	89	7.9	537	10 Q9SZ38	Q9sz38 arabidopsis
27	89	7.9	843	4 Q05707	Q05707 homo sapien
28	88.5	7.8	1253	6 Q97566	Q97566 canis famil
29	88.5	7.8	1703	11 Q9Z019	Q9z019 mus musculu
30	88	7.8	532	2 Q9LJ15	Q9llj5 streptomyce
31	88	7.8	704	10 Q9FF49	Q9ff49 arabidopsis
32	87.5	7.7	574	10 Q9FNF8	Q9fnf8 arabidopsis
33	87	7.7	689	10 Q9ZV10	Q9zv10 arabidopsis
34	86.5	7.6	487	2 Q9L3L1	Q9l3l1 salmoneilla
35	86.5	7.6	765	5 Q9U8J9	Q9u8j9 neospora ca
36	85.5	7.6	425	5 Q9GZF5	Q9gzf5 caenorhabdi
37	85.5	7.6	453	10 Q49285	Q49285 arabidopsis
38	85.5	7.6	1478	4 Q9BQ58	Q9bq58 homo sapien
39	85	7.5	3767	5 Q9UA13	Q9ua13 caenorhabdi
40	84.5	7.5	237	16 Q06731	Q06731 bacillus su
41	84.5	7.5	240	17 Q74080	Q74080 pyrococcus
42	84.5	7.5	433	10 Q9AX09	Q9ax09 oryza sativ
43	84.5	7.5	1731	4 Q9P230	Q9p230 homo sapien
44	84.5	7.5	3124	4 Q96L91	Q96l91 homo sapien
45	84	7.4	603	12 P89140	P89140 pseudorabie

## ALIGNMENTS

RESULT 1

Q96P02	ID	Q96P02	PRELIMINARY:	PRT:	368 AA.
AC	Q96P02;				
DT	01-DEC-2001	(Tremblrel. 19, Created)			
DT	01-DEC-2001	(Tremblrel. 19, Last sequence update)			
DT	01-DEC-2001	(Tremblrel. 19, Last annotation update)			
DE	ANTHRAX TOXIN RECEPTOR.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Bradley K.A., Mowbride J., Mowbride J., Collier R.J., Young J.A.T.;				
RT	"Identification of the Cellular Receptor for Anthrax Toxin.";				
RL	Nature 414:0-0(2001).				
DR	EMBL; AF421380; AAL26496.1; -.				
KW	Receptor.				
SQ	SEQUENCE	368 AA;	41157 MW;	8A87B13FFA7D8753	CRC64;

Query Match	85.4%;	Score	966;	DB	4;	Length	368;
Best Local Similarity	98.9%;	Pred. No.	7.6e-81;				
Matches	188;	Conservative	2;	Mismatches	0;	Indels	0;
Qy	1	MSFVFSFGTTLMLKLTDR	QIRQGLEELQKVLPGGDTYMHGFERASEQIYYENRQY	60			
Db	80	MSFVFSFGTTLMLKLTDR	QIRQGLEELQKVLPGGDTYMHGFERASEQIYYENRQY	139			
Qy	61	RTASVITATDGLHEDLFFYSEREANRSRDLGAIVYCVGVKDFNETOLARIADSKHVF	120				
Db	140	RTASVITATDGLHEDLFFYSEREANRSRDLGAIVYCVGVKDFNETOLARIADSKHVF	199				
Qy	121	PVNDGFQALQIIHILKSCIEILAAEPSTICAGESFQVYVVRGNGFRHARNDRVLCSE	180				
Db	200	PVNDGFQALQIIHILKSCIEILAAEPSTICAGESFQVYVVRGNGFRHARNDRVLCSE	259				
Qy	181	KINDSVTLK	190				

Fri Aug 9 10:56:56 2002

```
|||||||::
Db 260 KINDSVTLNE 269

RESULT 2
Q96NC7 PRELIMINARY; PRT; 245 AA.
AC Q96NC7;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CDNA FLJ31074 FIS, CLONE HSYRA2001476.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saico K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RP "NEDO human cDNA sequencing project.";
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK055336; BAB70976.1; -
SQ SEQUENCE 245 AA; 26111 MW; 61AE6E90A2EBEE06 CRC64;

Query Match 44.7%; Score 506; DB 4; Length 245;
Best Local Similarity 53.4%; Pred. No. 1.1e-38;
Matches 101; Conservative 37; Mismatches 49; Indels 2; Gaps 1;

QY 1 MSFVFSRGTMLKLTEDRQIQGLEELQKVLPGDGYMHEGFERASQIYYENRQY 60
D 3 LSFVFSQATIIPLTGDRKISGLGDLKRVSPVGETIHEGLKLANEQI--QKAGGL 60
QY 61 RTASVITIALDGEHEDLFYSEANRDLGAIVCYGVKDNFETQLARIADSKDHVF 120
D 61 KTSSIIITDGLDGLVPSYAEAKISRLGASVYCVGLDFEQALERIADSKSQVF 120
QY 121 PVNDGFALQGIHLSILKSKCIEILAAEPSTICAGESQVVRNGRHNVDRLVCSF 180
D 121 PVKGGFQALQGIHLSILKSKCIEILAAEPSTICAGESQVVRNGRHNVDRLVCSF 180
QY 181 KINDSVTLN 189
D 181 TVNEYTTS 189

RESULT 3
Q96EC6 PRELIMINARY; PRT; 97 AA.
AC Q96EC6;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE UNKNOWN (PROTEIN FOR IMAGE:4705862) (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Straussberg R.;
RA Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC012475; AAB12475.1; -
FT NON_TER 1
SQ SEQUENCE 97 AA; 10453 MW; 14F475F0B170E71A CRC64;

Query Match 13.4%; Score 151; DB 4; Length 97;
Best Local Similarity 87.9%; Pred. No. 1.6e-06;
Matches 29; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 158 FOVVVRNGRHNVDRLVCSFKINDSVTLN 190
D 1 FOVVVRNGRHNVDRLVCSFKINDSVTLN 33

RESULT 4
Q9BPQ8 PRELIMINARY; PRT; 1332 AA.
AC Q9BPQ8;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE INTEGRIN ALPHA HRI PRECURSOR.
GN HRTGAL.
OS Halocynthia roretzi (Sea squirt).
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
OC Stolidobranchia; Pyuridae; Halocynthia.
OX NCBI_TaxID=7729;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=HEMOCYTE;
RX MEDLINE=21103187; PubMed=11160215;
RA Miyazawa S., Azumi K., Nonaka M.;
RT "Cloning and characterization of integrin alpha subunits from the
RT solitary ascidian, Halocynthia roretzi.";
RL J. Immunol. 166:1710-1715(2001).
DR EMBL; AB048261; BAB21479.1; -
DR HSP; P11215; IABX.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF01839; FG-GAP; 5.
DR Pfam; PF00357; integrin_A; 1.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWFA; 1.
KW Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1332 INTEGRIN ALPHA HRI.
SQ SEQUENCE 1332 AA; 145852 MW; 0D9108D2B05CFFAE CRC64;

Query Match 10.9%; Score 123; DB 5; Length 1332;
Best Local Similarity 25.7%; Pred. No. 0.02;
Matches 49; Conservative 29; Mismatches 67; Indels 46; Gaps 7;

QY 15 KLTEDEQIRQGLEELQKVLPG-----GDTYMEHGFERASEQIYYENRQ 59
D 234 KLDIGKIVRVGVVQYSHYVEGSINKKQYITTEISIEGKFLDNFENAVDRI---QLQ 290
QY 60 YRT-----ASVITIALDGEHEDLFYSEANRDLGAIVCYV 99
D 291 YTTTGRALQKIVRDFDDAYIGNKQVLLLTLDGQAKDKLILP--NANRLNRKIATFV 348
QY 100 GVKDFNETQLARIA---DSKDHVPVNDGFALQGIHLSILKSKCIEILAAEPSTICAGE 156
D 349 GVGEYDISELKLIASGTSDTRVFTVD-FGEILDSIVKSLQTEIQSVLEGGKSAKTAG- 406
QY 157 SFQVVRNGNF 167
D 407 -YEMHGFENG 416

RESULT 5
Q9QYE7
```

QOQE7	PRELIMINARY;	PRT;	1161 AA.
AC	QOQE7;		
DT	01-MAY-2000 (Tremblrel. 13, Created)		
DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)		
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)		
DE	ALPHA D INTEGRIN.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.		
OX	NCBI_TaxID=10116;		
[1]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=SPRAGUE-DAWLEY;		
RA	O'Brien M.M., Vandervieren M., Kilgannon P.D., Dietsch G.,		
RA	Gallatin W.M.;		
RT	"Cloning of rat alpha D, a novel beta 2 integrin.";		
RL	Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.		
EMBL	AF021334; AAF21241.1; -		
HSP	P11215; IABX.		
DR	InterPro: IPR000413; Integrin_alpha.		
DR	InterPro: IPR002035; VWFA.		
DR	Pfam: PF01839; FG-GAP; 5.		
DR	Pfam: PF00357; Integrin_A; 1.		
DR	Pfam: PF00092; vwa; 1.		
DR	PRINTS: PR01185; INTEGRINA.		
DR	PRINTS: PR00453; VWFADOMAIN.		
DR	SMART: SM00191; Int_alpha; 4.		
DR	SMART: SM00327; VWA; 1.		
DR	PROSITE: PS00242; INTEGRIN_ALPHA; 1.		
DR	PROSITE: PS00234; VWFA; 1.		
QO	Integrin.		
SW	SEQUENCE		
QO	SEQUENCE	1161 AA; 126600 MW; 2258491A984A705E CRC64;	

Query Match	10.8%;	Score 122;	DB 11;	Length 1161;
Best Local Similarity	27.7%;	Pred. No. 0.021;		
Matches 52;	Conservative 34;	Mismatches 76;	Indels 28;	Gaps 10;
Qy	2	SFIVFSTRGTTMLKLT-----DREIQRQLEELQKVLPGGDTVMHGFERASQIYYEN	56	
		↓ : : : ↓ ↓ ↓ : : : ↓ ↓ ↓ : : : ↓ ↓ ↓ : : : ↓		
Db	190	SLMOYSNILKTHFTTEFKNILDQSLVDPIVQLQ-----GLTTATGIRTVMELFHSK	244	
Qy	57	RQGYRTA-SVIIATDGHEDLFYSE--REANRSRDIGAIVCVGVKD-FNE-TQLAR	111	
		: : : : : ↓ ↓ ↓ : : : : : ↓ ↓ ↓ : : : : : ↓		
Db	245	NGSRSAKKILLVTDGQKYRDPLEYSVPIAADKA---GIIRVAIGVDQAFQPTALKE	301	
Qy	112	I-----ADSKDHVPVNDGFOALGIIHSILKKSIEETLAAEPSTICAGESFQVVVRNG	166	
		↓ : : : ↓ ↓ ↓ : : : ↓ ↓ ↓ : : : ↓ ↓ ↓ : : : ↓		
Db	302	LNTTGSAPPQHVFKVGN-FAALKSIORLQEK-----IFAETGQSRSSSSFOHEMSOEG	356	
Qy	167	FRHARNVD	174	
		↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓		
Db	357	FSSALISD	364	

RESULT	6
ID	Q28984
PRT:	PRELIMINARY;
ID	Q28984
AC	Q28984;
DT	01-NOV-1996 (TtEMBLrel. 01, Created)
DT	01-FEB-1997 (TtEMBLrel. 02, Last sequence update)
DT	01-DEC-2001 (TtEMBLrel. 19, Last annotation update)
DE	COL1B (FRAGMENT).
GN	CD11B.
OS	Sus scrofa (Pig).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX	[1]
RN	SEQUENCE FROM N.A.
RP	Lee J.-K., Schook L.B., Rutherford M.S.;
RA	Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RL	EWBL; U40072; AAB16869.1; -.
DR	

[illegible][illegible]

Query Match 9.5%; Score 107.5; DB 4; Length 517;  
Best Local Similarity 26.0%; Pred. No. 0.15;  
Matches 60; Conservative 35; Mismatches 93; Indels 43; Gaps 12;  
Qv 12 TLMKLTDEQIIRGLEELQKVLDP--GGDTYMHGEGFRASEIYVENRGYRT--ASVII 67

Db	194	TLKKETKVEDII-----EAINTEFYGGSTNTGKAMTVREKIFVPK-GSRSNVFKVMI	247
Qy	68	ALTDCELHEDLFFYSEREANRSDLGAIVYGVKDFNETOLARIAD--SKDHVPVNDG	125
Db	248	LITDCK-SSDAF---RDPAILKRNDSVEIFAVGVKDAVDSELEAIAASPAAETHVFTVED-	302
Qy	126	FOALGITHSLIKKSCIEI---LAA-----EPSTICAGESFQVVRGNGFRHARNVD	175
Db	303	FDAFQRISEFTQSTICLIEQELAAIKKAYPPKDLSEVTSYGFKNWSPAGENVFS	362
Qy	176	VLCSFKI---NDSVTLTKSKLSQP-----WVSTSGFKEGNSHP	210
Db	363	YHITYKEAGDDVENTVWEPASSTSVLWLSKPTLYLVNVNVAEYEDGFSIP	413
RESULT	8		
070350			
-----			
CPYIYNADV.			
			PRT: 760 AA.

AC	070350;	
AD	01-AUG-1998 (TReMBLrel. 07, Created)	
AE	01-AUG-1999 (TReMBLrel. 11, Last sequence update)	
AF	01-DEC-2001 (TReMBLrel. 19, Last annotation update)	
AG	COMPLEMENT FACTOR C2 (COMPLEMENT COMPONENT 2) (WITHIN H-2S).	
AH	C2.	
AI	Mus musculus (Mouse).	
AM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AN	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AO	NCBI_TaxID=10090;	
AP	[1]	
AR	SEQUENCE FROM N.A.	
AS	STRAIN=129;	
AT	Rowen L., Madan A., Qin S., Hall J., Dahl T., James R., Dickhoff R.,	
AV	Schaffer T., Ratcliffe A., Abassi N., Loretz C., Lasky S., Hood L.;	
AW	"Sequence of the mouse MHC class III region.";	
AX	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.	
AY	[2]	
BA	SEQUENCE OF 291-760 FROM N.A.	
BB	Rowen L., Qin S., Lasky S.R., Loretz C., Dors M., Mahairas G.,	
BC	Hood L.;	
BD	Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.	
BE	[3]	
BF	SEQUENCE FROM N.A.	
BG	TISSUE=BREAST TUMOR;	
BH	Strausberg R.	
BI	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.	
BJ	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE	
BK	TRYPSIN FAMILY.	
BL	EMBL: AF109906; AAC84162.1; -.	
BM	EMBL: AF049850; AAC05284.1; -.	

```

DR DR HSSP; P00761; LEPT.
DR DR MGD; MGI:88226; C2.
DR DR InterPro; IPR001314; Chymotrypsin.
DR DR InterPro; IPR000436; Sushi_SCR_CCP.
DR DR InterPro; IPR001254; Trypsin.
DR DR InterPro; IPR002035; vWFA.
DR DR Pfam; PF00084; sushi; 2.
DR DR Pfam; PF00089; trypsin; 2.
DR DR Pfam; PF00092; vwa; 1.
DR DR PRINTS; PR00722; CHYMOTRYPSIN.
DR DR PRINTS; PR00453; VWFADOMAIN.
DR DR SMART; SM00032; CCP; 2.
DR DR SMART; SM00020; TRYP_SPC; 1.
DR DR SMART; SM00327; VWA; 1.
DR DR PROSITE; PS02440; TRYPsin_DOM; 2.
DR DR PROSITE; PS00134; TRYPsin_SER; 1.
DR DR PROSITE; PS00135; TRYPsin_SER; UNKNOWN_1.
DR DR PROSITE; PS02234; VWFA; 2.
DR DR PROSITE; PS02234; VWFA; 2.
DR DR Hydrolase; Serine protease.
DR DR SEQUENCE 760 AA; 84741 MW; 91c896a3edc7d448 CRC64;
DR DR

```

AC Q63001;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE INTEGRIN ALPHA-M (FRAGMENT).  
GN ITGAM.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WISTAR KYOTO;  
RA Nabika T., Ito T., Kitada H., Serikawa T., Mahimo T., Soubrier F.,  
RA Julier C., Masuda J., Yamori Y., Nara Y.;  
RT "Comparative mapping of a hypertension-related region on rat  
RT chromosome 1.";  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U59801; AAB03226.1; -.  
DR HSP; P11215; IABX.  
DR InterPro; IPR002035; VWFA.  
DR Pfam; PF00092; vwa; 1.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00327; VWA; 1.  
DR PROSITE; PS50234; VWFA; 1.  
FT NON\_TER 1 205  
FT NON\_TER 205 205  
SQ SEQUENCE 205 AA; 22922 MW; C8C2D9395008DA36 CRC64;

Query Match 9.2%; Score 104; DB 11; Length 205;  
Best Local Similarity 28.2%; Pred. No. 0.094;  
Matches 40; Conservative 24; Mismatches 64; Indels 14; Gaps 6;

Qy 37 GDTYMHGFERASEQIYYE-NRQGYRTASVITLTDGELHEDLFYFSERANRDLGAI 95  
Db 51 GRTKTASGIRKVVRELFQKINGARDNAKILVITDGEKFGDPLNY-EDVIPEAEAGII 109  
Qy 96 VYCVGVKD-FNETOLARIAD-----SKDHVFPVNDGFOALQGIHSLKSCIEILAAE 148  
Db 110 RYVIGVGNFNAFKPQSRRELDITASKPAGDHVFQV-DNEALNTIRNLQEK-----IFAIE 164  
Qy 149 PSTICAGESFQVVRGNGFRHA 170  
Db 165 GTQTGSTSSFEHMSQEGFSAA 186

RESULT 11  
Q9UGC3  
ID Q9UGC3 PRELIMINARY; PRT; 660 AA.  
AC Q9UGC3;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE DJ234P15.1 (COLLAGEN, TYPE XII, ALPHA 1) (FRAGMENT).  
GN COL12A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Smith M.;  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL080250; CABG5984.1; -.  
DR HSP; P11215; 1BHQ.  
DR InterPro; IPR003961; FN\_III.  
DR InterPro; IPR002035; VWFA.  
DR Pfam; PF00041; fn3; 2.  
DR Pfam; PF00092; vwa; 2.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00050; FN3; 2.  
DR SMART; SM00327; VWA; 2.

DR PROSITE; PS50234; VWFA; 2.  
KW Collagen.  
FT NON\_TER 660 660  
SQ SEQUENCE 660 AA; 72523 MW; 3E3A10A285ECA51 CRC64;

Query Match 9.1%; Score 103; DB 4; Length 660;  
Best Local Similarity 31.0%; Pred. No. 0.55;  
Matches 45; Conservative 22; Mismatches 58; Indels 20; Gaps 9;

Qy 12 TLMKLTEDREIQROGLEELQKLP--GGDTYMHGFERASEQIYYENRQGYRT--ASVII 67  
Db 522 TLKRFKVEDII-----EAINTFYRGSTNTGKAMTVYREKIEFVPSK-GSRSNVPKVM 575  
Qy 68 ALTDGELHEDLFYFSERANRDLGAIYCVGVKDFNETOLARIAD--SKDHVFPVNDG 125  
Db 576 LITDGR-SSDAF---RDPAILKRNDSVEIFAVGVKDAVRSELEAIASPPAETHVTVED- 630  
Qy 126 FOALQGIHSLKSCIEI---LAA 147  
Db 631 FDFQIRISFELTQSLCLRIEQELAA 655

RESULT 12  
Q9JI30  
ID Q9JI30 PRELIMINARY; PRT; 1151 AA.  
AC Q9JI30;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE INTEGRIN BETA 2 ALPHA SUBUNIT.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Fathallah D.M. Sr., Zerrila K. Jr.;  
RT Cloning of the rat CD11b cDNA sequence."  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF268593; AAF81280.1; -.  
DR HSP; P11215; IABX.  
DR InterPro; IPR000413; Integrin\_alpha.  
DR InterPro; IPR002035; VWFA.  
DR Pfam; PF01839; FG-GAP; 5.  
DR Pfam; PF00357; Integrin\_A; 1.  
DR Pfam; PF00092; vwa; 1.  
DR PRINTS; PR01185; INTEGRINA.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00191; Int\_alpha; 5.  
DR SMART; SM00327; VWA; 1.  
DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
DR PROSITE; PS50234; VWFA; 1.  
SQ SEQUENCE 1151 AA; 126943 MW; 8F785695D4074CA5 CRC64;

Query Match 9.1%; Score 103; DB 11; Length 1151;  
Best Local Similarity 28.2%; Pred. No. 1.2;  
Matches 40; Conservative 24; Mismatches 64; Indels 14; Gaps 6;

Qy 37 GDTYMHGFERASEQIYYE-NRQGYRTASVITLTDGELHEDLFYFSERANRDLGAI 95  
Db 223 GRTKTASGIRKVVRELFQKINGARDNAKILVITDGEKFGDPLNY-EDVIPEAEAGII 281  
Qy 96 VYCVGVKD-FNETOLARIAD-----SKDHVFPVNDGFOALQGIHSLKSCIEILAAE 148  
Db 282 RYVIGVGNFNAFKPQSRRELDITASKPAGDHVFQV-DNEALNTIRNLQEK-----IFAIE 336  
Qy 149 PSTICAGESFQVVRGNGFRHA 170  
Db 337 GTQTGSTSSFEHMSQEGFSAA 358

Fri Aug 9 10:56:56 2002

RA Cadieu E., Lelaure V., Galibert F., McDougall R.C., Rajandream M.A.,  
 RA Barrell B.G.;  
 RA Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; AL136235; CAB65811.1; -.  
 DR HSP; P17119; 3KAR.  
 DR InterPro; IPR001752; kinesin.  
 DR Pfam; PF002225; kinesin; 1.  
 DR PRINTS; PR00380; KINESINHEAVY.  
 DR SMART; SM00129; KISC; 1.  
 DR PROSITE; PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
 DR PROSITE; PS00067; KINESIN\_MOTOR\_DOMAIN2; 1.  
 DR ATP-binding; Coiled coil; Microtubules; Motor protein.  
 KW ATP-binding; Coiled coil; Microtubules; Motor protein.  
 SQ SEQUENCE 817 AA; 91043 MW; 804677851AF1A52D CRC64;

Query Match 8.5%; Score 96.5; DB 3; Length 817;  
 Best Local Similarity 25.1%; Pred. No. 2.9;  
 Matches 51; Conservative 33; Mismatches 88; Indels 31; Gaps 10;  
 QY 11 TLMKLTREDRIQROGLEELQKVLPGDGYMHGFERASEQIYYENRQGYRTASVIALT 70  
 Db 360 SAIQPLQENNSLKKQIQLELQREL-ASETVVYKENVKLSLDQ-QSANYVKLESTNRALEST 417  
 QY 71 DGELEDLFFYSE-----REANRSRDLGAIVYCVGVKDFNETQIARIADSKH--V 119  
 Db 418 IKTEEDVYTMKNIIELEGILKSANVERD-GLVEKLI AEETLRKHLNTIQELKGNIRV 476  
 QY 120 F-----PVNDGFQALQGIHISILKSCIEILAAEPSTTCAGESFQVVVGVNGFR-HARNV 173  
 Db 477 FCRVRPLGDGESAQIAFPDQNSEASTIEIVAQAP-----GSS-----LTGNGIKQYAFNF 527  
 QY 174 DRVLCSEKINDSV--TLKSLOS 194  
 Db 528 DRVFPSETTNEDEVNELSLOS 550

RESULT 15  
 Q96HX6 PRELIMINARY; PRT; 764 AA.  
 ID Q96HX6  
 AC Q96HX6;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE B-FACTOR, PROPERDIN.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=COLON ADENOCARCINOMA;  
 RA Strausberg R.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC007990; AA07990.1; -.  
 SQ SEQUENCE 764 AA; 85563 MW; 7574383E5FF7CC95 CRC64;

Query Match 8.4%; Score 95; DB 4; Length 764;  
 Best Local Similarity 20.3%; Pred. No. 3.7;  
 Matches 42; Conservative 32; Mismatches 69; Indels 64; Gaps 8;  
 QY 24 RQGLEELQKVLPGDGYMHGFERASEQIYYENRQGYRTASVIALTDG----- 72  
 Db 356 KKALQAVYSMSWPDVPPGWN-----RTRHVILMTDGLHNMGGDPITV 401  
 QY 73 --ELHEDLFFYSEREANKSRDLGAIVYCVG--VKDFNETQIARIADSKDHVPVNDGFOA 128  
 Db 402 IDEIRDLLYIGKDRKNPREYLDVYVFGVPLVQNVNTINALASKKDNQHVFKVKD-MEN 460  
 QY 129 LQGIHISILKSCIEILAAEPSTTCAGESFQVVVGVNGFRHARNVDVLCSEKINDSVTL 188  
 Db 461 LEDVFYQIMDES-----QSLSLC-----GMVWEHRKGTD----- 489

RA Cadieu E., Lelaure V., Galibert F., McDougall R.C., Rajandream M.A.,  
 RA Barrell B.G.;  
 RA Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; AL136235; CAB65811.1; -.  
 DR HSP; P17119; 3KAR.  
 DR InterPro; IPR001752; kinesin.  
 DR Pfam; PF002225; kinesin; 1.  
 DR PRINTS; PR00380; KINESINHEAVY.  
 DR SMART; SM00129; KISC; 1.  
 DR PROSITE; PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
 DR PROSITE; PS00067; KINESIN\_MOTOR\_DOMAIN2; 1.  
 DR ATP-binding; Coiled coil; Microtubules; Motor protein.  
 KW ATP-binding; Coiled coil; Microtubules; Motor protein.  
 SQ SEQUENCE 817 AA; 91043 MW; 804677851AF1A52D CRC64;

Query Match 8.8%; Score 99; DB 16; Length 698;  
 Best Local Similarity 27.0%; Pred. No. 1.4;  
 Matches 33; Conservative 28; Mismatches 45; Indels 16; Gaps 5;  
 QY 1 MSFVSTRGTILMKLTREDRIQROGLEELQKVLPGDGYMHGFERASEQIYYENRQGY 60  
 Db 550 LALIGFADKVKTLNLTEDREIEIFRAIDGLKKADVGTST-MSEPFSEA---YNILKDAY 604  
 QY 61 RTASVIALTDGELHEDLFFYSER-----EANKSRDLGAIVYCVGVKDFNETQIARIADSK 116  
 Db 605 GDCEVVV-LTDGQ-----WYKKDIMAENVKCKEYIEIAAIFGNKAKKDFLDKIATCE 657  
 QY 117 DH 118  
 Db 658 EN 659

RESULT 14  
 Q9US03 PRELIMINARY; PRT; 817 AA.  
 ID Q9US03  
 AC Q9US03;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE KINESIN-LIKE PROTEIN.  
 GN SPAC664.10.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972H-;



Qy 189 SKSLQSPWYSTSGFKEGNSH-PCLPA 214  
Db 490 --YHKQWQAKISVIRPSKGHESMCA 514

Search completed: August 9, 2002, 10:46:28  
Job time: 851 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:47:07 ; Search time 31.3 Seconds  
(without alignments)  
269.676 Million cell updates/sec

Title: US-09-970-076-10  
Perfect score: 1131  
Sequence: 1 MSFVSTRTTLMLKLTEDR.....STSGFKEGNSHPCLPARPHT 218

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result NO.	Score	Query Match %	Length	DB ID	Description
1	966	85.4	564	1 ATR_HUMAN	Q9h6x2 homo sapien
2	960	84.9	562	1 ATR_MOUSE	Q9cz52 mus musculus
3	373	33.0	386	1 CMG2_HUMAN	P58335 homo sapien
4	114	10.1	1162	1 ITAD_HUMAN	Q13349 homo sapien
5	114	10.1	3124	1 CA1C_CHICK	P13944 gallus gall
6	112	9.9	1153	1 ITAM_MOUSE	P03555 mus musculus
7	111	9.8	1152	1 ITAM_HUMAN	P11215 homo sapien
8	109	9.6	929	1 CA1C_NOTVI	Q91145 notophthalm
9	109	9.6	3067	1 CA1C_MOUSE	Q60847 mus musculus
10	106.5	9.4	760	1 CO2_MOUSE	P21180 mus musculus
11	105.5	9.3	3063	1 CA1C_HUMAN	Q99715 homo sapien
12	103	9.1	1163	1 ITAX_HUMAN	P20702 homo sapien
13	102.5	9.1	1888	1 CAME_CHICK	P32018 gallus gall
14	96.5	8.5	496	1 CAME_HUMAN	P21941 homo sapien
15	95	8.4	639	1 CA1C_RABIT	Q28902 oryctolagus
16	95	8.4	764	1 CFAB_HUMAN	P00751 homo sapien
17	94	8.3	452	1 MTN3_CHICK	Q42401 gallus gall
18	93.5	8.3	500	1 CANA_MOUSE	P51942 mus musculus
19	92	8.1	956	1 MTN2_HUMAN	Q00339 homo sapien
20	88	7.8	2944	1 CA17_HUMAN	Q02388 homo sapien
21	86.5	7.6	487	1 PHOQ_SALTY	P14147 salmonella
22	85.5	7.6	3176	1 YN36_HUMAN	P12111 homo sapien
23	85	7.5	3051	1 CA36_MOUSE	P34576 caenorhabdi
24	83.5	7.4	432	1 Y849_AQUAE	Q67016 aquifex aeo
25	83	7.3	956	1 MTN2_MOUSE	Q08746 mus musculus
26	82	7.3	761	1 CFAB_MOUSE	P04186 mus musculus
27	81.5	7.2	750	1 PBXP_STRPN	P14677 streptococc
28	81.5	7.2	897	1 DP01_HELPJ	Q9zje9 helicobacte
29	81	7.2	189	1 YN33_CAEEL	P34576 caenorhabdi
30	81	7.2	260	1 1433_ORYSA	Q06967 oryza sativ
31	81	7.2	265	1 1437_ARATH	Q96300 arabidopsis
32	80.5	7.1	796	1 PRH_ARATH	P48785 arabidopsis
33	80.5	7.1	3137	1 CA36_CHICK	P15989 gallus gall

34 79.5 7.0 493 1 CAME\_CHICK P05099 gallus gall  
35 79.5 7.0 619 1 OM70\_NEUCR P23231 neurospora  
36 79.5 7.0 935 1 MSH2\_MOUSE P43247 mus musculus  
37 79 7.0 262 1 143A\_HORVU P29305 hordeum vul  
38 78.5 6.9 1167 1 ITAE\_MOUSE Q60677 mus musculus  
39 78.5 6.9 1179 1 ITAE\_HUMAN P38570 homo sapien  
40 78.5 6.9 1436 1 DPO3\_STAAU Q53665 staphylococ  
41 78 6.9 259 1 1433\_LILLO Q9sp07 lilium long  
42 78 6.9 342 1 DHSY\_PRRHO O50105 pyrococcus  
43 77.5 6.9 673 1 TF2B\_METJA Q58192 methanococc  
44 77 6.8 1028 1 CA16\_HUMAN P12109 homo sapien  
45 76.5 6.8 179 1 ARF5\_CHICK P49702 gallus gall

#### ALIGNMENTS

```
RESULT 1
ATR_HUMAN
ID ATR_HUMAN STANDARD; PRT; 564 AA.
AC Q9H6X2; Q9NVP3;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Antrax toxin receptor precursor (Tumor endothelial marker 8).
GN ATR OR TEM8
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20407466; PubMed=10947988;
RA St Croix B., Rago C., Velculescu V., Traverso G., Romans K.E.,
RA Montgomery E., Lal A., Riggins G.J., Lengauer C., Vogelstein B.,
RA Kinzler K.W.;
RT "Genes expressed in human tumor endothelium.";
RL Science 289:1197-1202(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX PubMed=11700562;
RA Bradley K.A., Mogridge J., Mourez M., Collier R.J., Young J.A.T.;
RT "Identification of the cellular receptor for anthrax toxin.";
RL Nature 414:225-229(2001).
RN [3]
RP SEQUENCE OF 184-564 FROM N.A. (ISOFORM 1), AND SEQUENCE FROM N.A.
(ISOFORM 3).
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 4).
RT TISSUE=Kidney;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Cellular role is not yet known.
CC -!- SUBUNIT: Binds to the protective antigen (PA) of Bacillus anthracis.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -!- ALTERNATIVE PRODUCTS: 4 isoforms; 1 (shown here), 2, 3 and 4; seem to be produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Highly expressed in tumor endothelial cells but not in normal endothelial cells.
CC -!- DOMAIN: Binding to PA seems to be effected through the VWA domain.
CC -!- SIMILARITY: BELONGS TO THE ATR FAMILY.
CC -!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way
```





CC LINES AND SUBSETS OF PERIPHERAL BLOOD LEUKOCYTES AND STRONGLY ON  
CC TISSUE-SPECIALIZED CELLS, INCLUDING MACROPHAGES FOAM CELLS WITHIN  
CC ATHEROSCLEROTIC PLAQUES, AND ON SPLENIC RED PULP MACROPHAGES.  
CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS  
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U37028; AAB38547.1; -  
CC EMBL; U40274; AAB60634.1; -  
CC EMBL; U40275; AAB60635.1; -  
CC EMBL; U40276; AAB60636.1; -  
CC EMBL; U40277; AAB60637.1; -  
CC EMBL; U40278; AAB60638.1; JOINED.  
CC EMBL; AF187881; AAF62875.1; -  
CC HSP; P11215; 1A8X.  
CC MIM; 602453; -  
CC InterPro; IPR000413; Integrin\_alpha.  
CC InterPro; IPR002035; VWFA.  
CC Pfam; PF01839; FG-GAP 5.  
CC Pfam; PF03357; Integrin\_A; 1.  
CC Pfam; PF00092; vwa; 1.  
CC PRINTS; PR01185; INTEGRINA.  
CC PRINTS; PR00453; VWFADOMAIN.  
CC SMART; SM00191; Int\_alpha; 4.  
CC SMART; SM00327; VWA; 1.  
CC PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
CC PROSITE; PS00234; VWFA; 1.  
CC Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
CC Signal; Repeat; Calcium;  
CC Magnesium.  
CC SIGNAL 1 17  
CC CHAIN 18 1162  
CC DOMAIN 18 1100  
CC TRANSMEM 1101 1124  
CC DOMAIN 1125 1162  
CC REPEAT 32 85  
CC REPEAT ? ?  
CC DOMAIN 150 332  
CC REPEAT 350 400  
CC REPEAT 401 452  
CC REPEAT 454 516  
CC REPEAT 518 576  
CC REPEAT 581 633  
CC CA\_BIND 465 473  
CC CA\_BIND 530 538  
CC CA\_BIND 593 601  
CC SITE 1127 1131  
CC DISULFID 67 74  
CC DISULFID 106 124  
CC DISULFID 655 710  
CC DISULFID 769 775  
CC DISULFID 846 861  
CC DISULFID 994 1018  
CC DISULFID 1023 1028  
CC CARBOHYD 59 59  
CC CARBOHYD 87 87  
CC CARBOHYD 99 99  
CC CARBOHYD 391 391  
CC CARBOHYD 691 691  
CC CARBOHYD 733 733  
CC CARBOHYD 873 873  
CC CARBOHYD 957 957

FT CARBOHYD 1046 1046 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CONFLICT 500 500 MISSING (IN REF. 2).  
FT CONFLICT 515 518 GHEW -> ATP (IN REF. 2).  
FT CONFLICT 825 825 L -> V (IN REF. 2).  
FT CONFLICT 984 984 V -> A (IN REF. 2).  
SQ SEQUENCE 1162 AA; 126885 MW; F296A1A3455D77D CRC64;  
  
Query Match 10.1%; Score 114; DB 1; Length 1162;  
Best Local Similarity 27.3%; Pred. No. 0.043;  
Matches 48; Conservative 27; Mismatches 83; Indels 18; Gaps 6;  
  
QY 11 TTMKLTREDRIQGLLEELQKLP-----GGDTMHGFEFRASQIYYENRQGYRTA-SV 65  
DB 193 SNLLKHITFTQTSPPSQSLVDPVQLKGLTFTATGILTVTQLEFHHKNGAKRSKKI 252  
QY 66 ITALTDELHEDLFYSEERANRDLGAIVVCVKDFNETQLAR-----TADSKDH 118  
DB 253 LIVITDGGQYKDPLEYSD-VIPOAEKAGLIYAIGVGHAFQGTARQELNTISSAPQDH 311  
QY 119 VFPVNDGFOALQGIHSLIKSCIEILAAEPSTICAGESFQVYVVGNGFRHARNVD 174  
DB 312 VFKV-DNFAALGSIQKLOEK----IYAVEGTQSRASSFOHMSQEGFSTALIMD 362  
  
RESULT 5  
CALC\_CHICK STANDARD; PRT; 3124 AA.  
ID CALC\_CHICK STANDARD; PRT; 3124 AA.  
AC P13944; Q04509;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Collagen alpha 1(XII) chain precursor (Fibrochimerin).  
GN COL12A1.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WHITE LEGHORN;  
RX MEDLINE=92011862; PubMed=1918137; Shinomura T., Tanaka H.,  
RA Yamagata M., Yamada K.M., Yamada S.S., Shinomura T., Tanaka H.,  
RA Nishida Y., Obara M., Kimata K.;  
RT "The complete primary structure of type XII collagen shows a chimeric  
RT molecule with reiterated fibronectin type III motifs, von Willebrand  
RT factor A motifs, a domain homologous to a noncollagenous region of  
RT type IX collagen, and short collagenous domains with an Arg-Gly-Asp  
RT site.";  
RL J. Cell Biol. 115:209-221(1991).  
RN [2]  
RP SEQUENCE OF 2456-3124 FROM N.A., AND SEQUENCE OF 2772-2794 AND  
RP 2846-2873  
RX MEDLINE=90062079; PubMed=2584192;  
RA Gordon M.K., Gerecke D.R., Dublet B., van der Rest M., Olsen B.R.;  
RT "Type XII collagen. A large multidomain molecule with partial  
RT homology to type IX collagen.";  
RL J. Biol. Chem. 264:19772-19778(1989).  
RN [3]  
RP SEQUENCE OF 2960-3076 FROM N.A.  
RX MEDLINE=87317590; PubMed=3476925;  
RA Gordon M.K., Gerecke D.R., Olsen B.R.;  
RT "Type XII collagen: distinct extracellular matrix component  
RT discovered by cDNA cloning.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:6040-6044(1987).  
RN [4]  
RP SEQUENCE OF 1-1283 FROM N.A. (SHORT FORM), AND ALTERNATIVE SPLICING.  
RC TISSUE-EMBRIO;  
RX MEDLINE=93042014; PubMed=1420368;  
RA Trueb J., Trueb B.;  
RT "The two splice variants of collagen XII share a common 5' end.";  
RL Biochim. Biophys. Acta 1171:97-98(1992).







RESULT 7  
ID ITAM\_HUMAN STANDARD; PRT; 1152 AA.  
AC P11215;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Integrin alpha-M precursor (Cell surface glycoprotein MAC-1 alpha subunit) (CR-3 alpha chain) (CD11b) (Leukocyte adhesion receptor M01)  
DE Neutrophil adherence receptor).  
GN ITGAM OR CR3A OR CD11B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88315033; PubMed=2457584;  
RA Corbi A.L., Kishimoto T.K., Miller L.J., Springer T.A.;  
RT "The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD11b) alpha subunit. Cloning, primary structure, and relation to the integrins, von Willebrand factor and factor B.";  
RL J. Biol. Chem. 263:12403-12411(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88190151; PubMed=2833753;  
RA Arnaout M.A., Remold-O'Donnell E., Pierce M.W., Harris P., Tenen D.G.;  
RT "Molecular cloning of the alpha subunit of human and guinea pig leukocyte adhesion glycoprotein Mol: chromosomal localization and homology to the alpha subunits of integrins.";  
RL Proc. Natl. Acad. Sci. U.S.A. 85:2776-2780(1988).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88257215; PubMed=2454931;  
RA Arnaout M.A., Gupta S.K., Pierce M.W., Tenen D.G.;  
RT "Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor Mol (complement receptor type 3).";  
RL J. Cell Biol. 106:2153-2158(1988).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93123748; PubMed=8419480;  
RA Fleming J.C., Pahl H.L., Gonzalez D.A., Smith T.F., Tenen D.G.;  
RT "Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-integrin gene family demonstrate remarkable conservation of genomic organization and suggest early diversification during evolution.";  
RL J. Immunol. 150:480-490(1993).  
RN [5]  
RP SEQUENCE OF 9-1153 FROM N.A.  
RX MEDLINE=8909893; PubMed=2563162;  
RA Hickstein D.D., Hickey M.J., Ozols J., Baker D.M., Back A.L., Roth G.J.;  
RT "cDNA sequence for the alpha M subunit of the human neutrophil adherence receptor indicates homology to integrin alpha subunits.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:257-261(1989).  
RN [6]  
RP SEQUENCE OF 1-9 FROM N.A.  
RX MEDLINE=92073318; PubMed=1683702;  
RA Shelley C.S., Arnaout M.A.;  
RT "The promoter of the CD11b gene directs myeloid-specific and developmentally regulated expression.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:10525-10529(1991).  
RN [7]  
RP SEQUENCE OF 1-9 FROM N.A.  
RC TISSUE=Blood;  
RX MEDLINE=92144986; PubMed=1346576;  
RA Pahl H.L., Rosmarin A.G., Tenen D.G.;  
RT "Characterization of the myeloid-specific CD11b promoter.";  
RL Blood 79:865-870(1992).  
RN [8]  
RP SEQUENCE OF 17-31.  
RX MEDLINE=87076671; PubMed=3539202;  
RA Pierce M.W., Remold-O'Donnell E., Todd R.F. III, Arnaout M.A.;  
RT "N-terminal sequence of human leukocyte glycoprotein Mol: conservation across species and homology to platelet IIB/IIIA.";  
Biochim. Biophys. Acta 874:368-371(1986).  
RN [9]  
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 148-331.  
RX MEDLINE=95171458; PubMed=7867070;  
RL Lee J.O., Rieu P., Arnaout M.A., Liddington R.;  
RT "Crystal structure of the A domain from the alpha subunit of integrin CR3 (CD11b/CD18).";  
RL Cell 80:631-638(1995).  
RN [10]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 148-334.  
RX MEDLINE=96363671; PubMed=8747460;  
RL Lee J.O., Bankston L.A., Arnaout M.A., Liddington R.C.;  
RT "Two conformations of the integrin A-domain (I-domain): a pathway for activation?";  
RL Structure 3:1333-1340(1995).  
RN [11]  
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 148-337.  
RX MEDLINE=98362595; PubMed=9687375;  
RL Baldwin E.T., Sarver R.W., Bryant G.L. Jr., Curry K.A., Fairbanks M.B., Finzel B.C., Garlick R.L., Heinrichson R.L., Horton N.C., Kelley L.L., Mildner A.M., Moon J.B., Mott J.E., Mutchler V.T., Tomich C.S., Watenpaugh K.D., Wiley V.H.;  
RT "Cation binding to the integrin CD11b I domain and activation model assessment.";  
RL Structure 6:923-935(1998).  
RN [12]  
RP 3D-STRUCTURE MODELING OF 17-616.  
RX MEDLINE=98226734; PubMed=9560195;  
RL Oxvig C., Springer T.A.;  
RT "Experimental support for a beta-propeller domain in integrin alpha-subunits and a calcium binding site on its lower surface.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:4870-4875(1998).  
CC -!- FUNCTION: INTEGRIN ALPHA-M/BETA-2 IS IMPLICATED IN VARIOUS ADHESIVE INTERACTIONS OF MONOCYTES, MACROPHAGES AND GRANULOCYTES AS WELL AS IN MEDIATING THE UPTAKE OF COMPLEMENT-COATED PARTICLES. IT IS IDENTICAL WITH CR-3, THE RECEPTOR FOR THE IC3B FRAGMENT OF THE THIRD COMPLEMENT COMPONENT. IT PROBABLY RECOGNIZES THE R-G-D PEPTIDE IN C3b. INTEGRIN ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR FIBRINOGEN, FACTOR X AND ICAM1. IT RECOGNIZES P1 AND P2 PEPTIDES OF FIBRINOGEN GAMMA CHAIN.  
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-M ASSOCIATES WITH BETA-2.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND GRANULOCYTES.  
CC -!- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
CC -!- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
CC -!- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
CC -!- DATABASE: NAME=PROW; NOTE=CD guide Cdl1b entry;  
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cdl1b.htm".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; J03925; AAA59544.1; -;  
DR EMBL; M18044; AAA59491.1; -;  
DR EMBL; J04145; AAA59903.1; -;  
DR EMBL; S52227; AAB24821.1; -;  
DR EMBL; S52152; AAB24821.1; JOINED.  
DR EMBL; S52153; AAB24821.1; JOINED.  
DR EMBL; S52154; AAB24821.1; JOINED.  
DR EMBL; S52155; AAB24821.1; JOINED.  
DR EMBL; S52157; AAB24821.1; JOINED.  
DR EMBL; S52159; AAB24821.1; JOINED.

RESULT	8	CAIC_NOTVI	STANDARD;	PRT;	929 AA.
EMBL; S52161;	DR	AB24821.1;	JOINED.		
EMBL; S52164;	DR	AB24821.1;	JOINED.		
EMBL; S52165;	DR	AB24821.1;	JOINED.		
EMBL; S52167;	DR	AB24821.1;	JOINED.		
EMBL; S52169;	DR	AB24821.1;	JOINED.		
EMBL; S52170;	DR	AB24821.1;	JOINED.		
EMBL; S52173;	DR	AB24821.1;	JOINED.		
EMBL; S52174;	DR	AB24821.1;	JOINED.		
EMBL; S52180;	DR	AB24821.1;	JOINED.		
EMBL; S52181;	DR	AB24821.1;	JOINED.		
EMBL; S52184;	DR	AB24821.1;	JOINED.		
EMBL; S52189;	DR	AB24821.1;	JOINED.		
EMBL; S52191;	DR	AB24821.1;	JOINED.		
EMBL; S52192;	DR	AB24821.1;	JOINED.		
EMBL; S52203;	DR	AB24821.1;	JOINED.		
EMBL; S52213;	DR	AB24821.1;	JOINED.		
EMBL; S52216;	DR	AB24821.1;	JOINED.		
EMBL; S52219;	DR	AB24821.1;	JOINED.		
EMBL; S52220;	DR	AB24821.1;	JOINED.		
EMBL; S52221;	DR	AB24821.1;	JOINED.		
EMBL; S52222;	DR	AB24821.1;	JOINED.		
EMBL; S52226;	DR	AB24821.1;	JOINED.		
EMBL; M76724;	DR	AA58410.1;	-.		
EMBL; M84477;	DR	AA51960.1;	-.		
PIR; A31108;	DR	RWHULB.			
PIR; A26091;	DR	A26091.			
PDB; 1A8X;	DR	17-JUN-98.			
PDB; 1BHO;	DR	18-NOV-98.			
PDB; 1BHQ;	DR	18-NOV-98.			
PDB; 1LDN;	DR	23-NOV-98.			
PDB; 1LDO;	DR	01-AUG-96.			
PDB; 1JLM;	DR	11-JAN-97.			
CarbBank; CCSD:33581;	DR	-.			
CarbBank; CCSD:33582;	DR	-.			
CarbBank; CCSD:33584;	DR	-.			
MIM; 120980;	DR	-.			
InterPro; IPR000413;	DR	Integrin_alpha.			
InterPro; IPR002035;	DR	VWFA.			
Pfam; PF01839;	DR	FG-GAP; 5.			
Pfam; PF00357;	DR	Integrin_A; 1.			
Pfam; PF00092;	DR	vwa; 1.			
PRINTS; PR01185;	DR	INTEGRINA.			
PRINTS; PR00453;	DR	VWFADOMAIN.			
SMART; SM00191;	DR	Int_alpha; 4.			
SMART; SM00327;	DR	VWA_1			
PROSITE; PS00242;	DR	INTEGRIN_ALPHA; 1.			
PROSITE; PS0234;	DR	VWFA; 1.			
InterPro; IPR003961;	DR	FN_III.			
InterPro; IPR002035;	DR	VWFA.			
Pfam; PF00041;	DR	fn3; 7.			
Pfam; PF00092;	DR	vwa; 1.			
SMART; SM00060;	DR	FN3; 5.			
SMART; SM00327;	DR	VWA; 1.			
PROSITE; PS0234;	DR	VWFA; 2.			
Extracellular matrix;	DR	Connective tissue; Repeat; Cell adhesion;			
Collagen; Glycoprotein.	DR				
NON_TER	DR	1			
DOMAIN	DR	<1			
DOMAIN	DR	63			
DOMAIN	DR	154			
DOMAIN	DR	155			
DOMAIN	DR	245			
DOMAIN	DR	338			
DOMAIN	DR	246			
DOMAIN	DR	339			
DOMAIN	DR	432			
DOMAIN	DR	433			
DOMAIN	DR	519			
DOMAIN	DR	49			
DOMAIN	DR	154			
DOMAIN	DR	245			
DOMAIN	DR	338			
DOMAIN	DR	246			
DOMAIN	DR	339			
DOMAIN	DR	432			





```
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PROSITE: PS02334; VWFA; 1.
KW Complement pathway; Plasma; Glycoprotein; Hydrolase; Serine protease;
KW Signal; Repeat; Sushi; Alternative splicing.
FT SIGNAL 1 18
FT CHAIN 19 760
FT CHAIN 19 760
FT CHAIN 251 760
FT CHAIN 251 760
FT DOMAIN 21 90
FT DOMAIN 93 150
FT DOMAIN 155 211
FT DOMAIN 261 459
FT DOMAIN 473 760
FT DOMAIN 514 514
FT ACT_SITE 514 514
FT ACT_SITE 570 570
FT ACT_SITE 689 689
FT ACT_SITE 689 689
FT DISULFID 22 62
FT DISULFID 49 89
FT DISULFID 94 136
FT DISULFID 122 149
FT DISULFID 156 197
FT DISULFID 182 210
FT CARBOHYD 27 27
FT CARBOHYD 32 32
FT CARBOHYD 117 117
FT CARBOHYD 297 297
FT CARBOHYD 340 340
FT CARBOHYD 474 474
FT CARBOHYD 478 478
FT CARBOHYD 663 663
FT VARSPIC 606 612
SQ SEQUENCE 760 AA; 84736 MW; D7DE9DEF4DBA48D9 CRC64;

Query Match 9.4%; Score 106.5; DB 1; Length 760;
Best Local Similarity 24.3%; Pred. No. 0.12;
Matches 57; Conservative 38; Mismatches 77; Indels 63; Gaps 12;

QY 1 MSFIVFSTRGTTLMKLTEDRQIGLELOKVLPGGDYVMEGPRASEQIYVE-----55
DB 300 VAIIFASQPKTMSILTSRSQ-----DVTEVITSLDSASYKDHNATGANTYVLIIRV 353

QY 56 -----NRQGYRTAS-----VIAITDGLHEDLFFYSEREANRSDLGAI-----95
DB 354 YSNMQTQMDRLGMETSASKEIRHTIILLTDGK--SNMGDSPPKAVTRIELLSIEQNRDD 411

QY 96 ---VYCVGV-----KDFNETQLARIADSKDHVPVNDGFOALQGIHSILKKSCIEIL 145
DB 412 YLDIYAIGVGKLDVDVKELNE--LGSKDGGERHAFILODA-KALQQIFEHMLDVSKL---465

QY 146 AAEPSTICAGESFQVVVRCNGFRHARNVRV--LCSEFKINDSVTSLSKSQS-PWV 197
DB 466 ---TDTCIG-----VGNMSANASDQERTPWQVTFKPKSKETCGSLISDQWV 509

RESULT 11
ID CALC_HUMAN STANDARD; PRT; 3063 AA.
AC Q99715; Q99716;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(XII) chain precursor.
GN COL12A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (LONG AND SHORT FORM).
RX MEDLINE=97288521; PubMed=9143499;
RA Gerecke D.R., Olson P.F., Koch M., Knoll J.H.M., Taylor R.,
RA Hudson D.L., Champlaud M.-F., Olsen B.R., Burgeson R.E.;
```

```
RT "Complete primary structure of two splice variants of collagen XII,
RT and assignment of alpha 1(XII) collagen (COL12A1), alpha 1(XI)
RT collagen (COL9A1), and alpha 1(XI) collagen (COL9A1) to human
RT chromosome 6q12-q13."
RL Genomics 41:236-242(1997).
CC -!- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-
CC CONTAINING FIBRILS, THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE
CC SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE
CC LOCALIZED IN THE PERIFIBRILLAR MATRIX (BY SIMILARITY).
CC -!- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 kDa OF
CC NONTRIPLE-HELICAL SEQUENCES.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE FINAL TISSUE
CC FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OF EITHER THE LONGER
CC OR THE SHORTER ISOFORM OR ANY COMBINATION OF LONG AND SHORT
CC ISOFORM CHAINS. ONLY THE LONG VARIANT IS A PROTEOLYCAN.
CC -!- TISSUE SPECIFICITY: FOUND IN COLLAGEN I-CONTAINING TISSUES: BOTH
CC SHORT AND LONG ISOFORMS APPEAR IN AMNION, CHORION, SKELETAL
CC MUSCLE, SMALL INTESTINE, AND IN CELL CULTURE OF DERMAL
CC FIBROBLASTS, KERATINOCYTES, AND ENDOTHELIAL CELLS. ONLY THE SHORT
CC ISOFORM IS FOUND IN LUNG, PLACENTA, KIDNEY, AND A SQUAMOUS CELL
CC CARCINOMA CELL LINE.
CC -!- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT
CC EACH END (BY SIMILARITY).
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY
CC SIMILARITY).
CC -!- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -!- SIMILARITY: CONTAINS 4 VWFA DOMAINS.
CC -!- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U73778; AAC51244.1; -.
DR EMBL; U73779; AAD40483.1; -.
DR HSSP; P02751; 1TTF.
DR MIM; 120320; -.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003962; FNIII_repeat.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF01391; Collagen; 4.
DR Pfam; PF00041; fn3; 18.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00092; vwa; 4.
DR PRINTS; PR00014; FNTYPEIII.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00060; FN3; 17.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00327; VWA; 4.
DR PROSITE; PS02334; VWFA; 4.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.
FT SIGNAL 1 24
FT CHAIN 25 3063
FT DOMAIN 25 114
FT DOMAIN 140 316
FT DOMAIN 333 426
FT DOMAIN 440 616
FT DOMAIN 630 721
FT DOMAIN 722 812
FT DOMAIN 813 903
FT DOMAIN 904 998
FT DOMAIN 999 1085
```

FT	DOMAIN	1086	1178	FIBRONECTIN TYPE-III 8.	AC	P20702;
FT	DOMAIN	1199	1371	WFPA 3.	DT	01-FEB-1991 (Rel. 17, Created)
FT	DOMAIN	1384	1473	FIBRONECTIN TYPE-III 9.	DT	01-FEB-1991 (Rel. 17, Last sequence update)
FT	DOMAIN	1474	1564	FIBRONECTIN TYPE-III 10.	DT	01-MAR-2002 (Rel. 41, Last annotation update)
FT	DOMAIN	1565	1652	FIBRONECTIN TYPE-III 11.	DE	Integrin alpha-X precursor (Leukocyte adhesion glycoprotein p150,95
FT	DOMAIN	1654	1751	FIBRONECTIN TYPE-III 12.	DE	alpha chain) (Leukocyte adhesion receptor p150,95) (CD11c) (Leu M5).
FT	DOMAIN	1752	1842	FIBRONECTIN TYPE-III 13.	GN	ITGAX OR CD11C.
FT	DOMAIN	1843	1932	FIBRONECTIN TYPE-III 14.	OS	Homo sapiens (Human).
FT	DOMAIN	1933	2023	FIBRONECTIN TYPE-III 15.	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
FT	DOMAIN	2024	2114	FIBRONECTIN TYPE-III 16.	OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
FT	DOMAIN	2115	2202	FIBRONECTIN TYPE-III 17.	OX	NCBI_TaxID=9606;
FT	DOMAIN	2203	2291	FIBRONECTIN TYPE-III 18.	[1]	SEQUENCE FROM N.A.
FT	DOMAIN	2323	2496	WFPA 4.	RP	MEDLINE=88166645; PubMed=3327687;
FT	DOMAIN	2451	2746	NONHELICAL REGION (NC3).	RX	Corbi A.L., Miller L.J., O'Connor K., Larson R.S., Springer T.A.;
FT	DOMAIN	2747	2898	TRIPLE-HELICAL REGION (COL2)	RA	"CDNA cloning and complete primary structure of the alpha subunit of
FT	DOMAIN	2899	2941	WITH 1 IMPERFECTION.	RT	a leukocyte adhesion glycoprotein, p150,95.";
FT	DOMAIN	2942	3044	NONHELICAL REGION (NC2).	RL	EMBO J. 6:4023-4028(1987).
FT	DOMAIN	3045	3063	TRIPLE-HELICAL REGION (COL1)	RN	[2]
FT	DOMAIN	798	798	WITH 2 IMPERFECTIONS.	RP	SEQUENCE FROM N.A.
FT	DOMAIN	889	889	NONHELICAL REGION (NC1).	RX	MEDLINE=90153906; PubMed=2303426;
FT	BINDING	891	981	TO CHONDROITIN SULFATE (POTENTIAL).	RA	Corbi A.L., Garcia-Aguilar J., Springer T.A.;
FT	BINDING	891	981	TO CHONDROITIN SULFATE (POTENTIAL).	RT	"Genomic structure of an integrin alpha subunit, the leukocyte
FT	SITE	862	864	CELL ATTACHMENT SITE (POTENTIAL).	RL	p150,95 molecule.";
FT	SITE	2779	2781	CELL ATTACHMENT SITE (POTENTIAL).	RN	J. Biol. Chem. 265:2782-2788(1990).
FT	SITE	2895	2897	CELL ATTACHMENT SITE (POTENTIAL).	[3]	ERRATUM.
FT	MOD_RES	2944	2944	HYDROXYLATION (BY SIMILARITY).	RP	Corbi A.L., Garcia-Aguilar J., Springer T.A.;
FT	MOD_RES	2947	2947	HYDROXYLATION (BY SIMILARITY).	RA	J. Biol. Chem. 265:12750-12751(1990).
FT	MOD_RES	2950	2950	HYDROXYLATION (BY SIMILARITY).	[4]	SEQUENCE OF 20-43.
FT	MOD_RES	2959	2959	HYDROXYLATION (BY SIMILARITY).	RP	MEDLINE=87167596; PubMed=3549901;
FT	MOD_RES	2965	2965	HYDROXYLATION (BY SIMILARITY).	RX	Miller L.J., Wiebe W., Springer T.A.;
FT	MOD_RES	2968	2968	HYDROXYLATION (BY SIMILARITY).	RA	"Purification and alpha subunit N-terminal sequences of human Mac-1
FT	MOD_RES	2971	2971	HYDROXYLATION (BY SIMILARITY).	RT	and p150,95 leukocyte adhesion proteins.";
FT	MOD_RES	2983	2983	HYDROXYLATION (BY SIMILARITY).	RL	J. Immunol. 138:2381-2383(1987).
FT	MOD_RES	3000	3000	HYDROXYLATION (BY SIMILARITY).	CC	FUNCTION: INTEGRIN ALPHA-X/BETA-2 IS A RECEPTOR FOR FIBRINOGEN. IT
FT	MOD_RES	3003	3003	HYDROXYLATION (BY SIMILARITY).	CC	RECOGNIZES THE SEQUENCE G-P-R IN FIBRINOGEN. IT MEDIATES CELL-CELL
FT	MOD_RES	3014	3014	HYDROXYLATION (BY SIMILARITY).	CC	INTERACTION DURING INFLAMMATORY RESPONSES. IT IS ESPECIALLY
FT	MOD_RES	3023	3023	HYDROXYLATION (BY SIMILARITY).	CC	IMPORTANT IN MONOCYTE ADHESION AND CHEMOTAXIS.
FT	MOD_RES	3026	3026	HYDROXYLATION (BY SIMILARITY).	CC	SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-X
FT	MOD_RES	3029	3029	HYDROXYLATION (BY SIMILARITY).	CC	ASSOCIATES WITH BETA-2.
FT	CARBOHYD	700	700	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	CELLULAR LOCATION: Type I membrane protein.
FT	CARBOHYD	1763	1763	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
FT	CARBOHYD	2206	2206	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
FT	CARBOHYD	2528	2528	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	SIMILARITY: CONTAINS 1 WFPA DOMAIN.
FT	CARBOHYD	2679	2679	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
FT	CARBOHYD	2679	2679	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC	DATABASE: NAME=PROW; NOTE=CD guide CDLIC entry;
FT	VARSPPLIC	25	1188	MISSING (IN SHORT ISOFORM).	CC	WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cdlic.htm".
SQ	SEQUENCE	3063 AA;	333189 MW;	75FEA78FA8E48293 CRC64;	CC	-----
Query Match					CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
Best Local Similarity					CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
Matches 60; Conservative					CC	the European Bioinformatics Institute. There are no restrictions on its
Qy	12	TLMLKLTEDRQIRGLELQKVL	-----GQDVMHEGFERASEQIYVENRQGYRT--ASVII 67		CC	use by non-profit institutions as long as its content is in no way
Db	493	TLKPKTKVEDII-----EAINTPFYRGSTNGTKAMTYVREKIFVPSK-GSRSNVPRKVI 546			CC	modified and this statement is not removed. Usage by and for commercial
Qy	68	ALTDGELHEDLFFYSERANRSDLGAIYCVGVKDFNETQLARIAD--SKDHVFPVNDG 125			CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
Db	547	LITDGK--SSDAF---RDPAILRNSDVEIFAVGVKDAVRSELEAIAASPPAETHVFTYED- 601			CC	or send an email to license@isb-sib.ch).
Qy	126	FOALQGIHSLIKSCIEI---LAA-----EPSTICAGESFOVVVRGNGFRHARNVDR 175			CC	-----
Db	602	FDFAQRISFELTQSCLERIEQLAAIKKAYVPRDLSFSEVTSYGKFNWSPAGENVFS 661			CC	EMBL; M61695; AAA59180.1;
Qy	176	VLCSEFKI---NDVSTLSKLSQSP-----WYSSSTSGFKEGNSHP 210			CC	EMBL; Y00093; CAA68283.1;
Db	662	YHITYKAGDDEVTVVEPASSTSVVLSLKPETILYLVNVTAEVEDGFSIP 712			CC	EMBL; M29165; -; NOT_ANNOTATED_CDS.
RESULT 12					CC	EMBL; M29487; AAA51620.1; ALT_SEQ.
ITAX_HUMAN					CC	EMBL; M29482; AAA51620.1; JOINED.
ID	ITAX_HUMAN	STANDARD;	PRT; 1163 AA.		CC	EMBL; M29483; AAA51620.1; JOINED.
					CC	EMBL; M29484; AAA51620.1; JOINED.
					CC	EMBL; M29485; AAA51620.1; JOINED.
					CC	EMBL; M29486; AAA51620.1; JOINED.
					CC	PIR; A36584; RWHUIC.
					CC	HSPP; P11215; IABX.
					CC	CarbBank; CCSD:33581; -.

```
DR CarbBank; CCSD:33582; -
DR CarbBank; CCSD:33584; -
DR MW; 151510; -
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF01839; FG-GAP; 5.
DR Pfam; PF00357; Integrin_A; 1.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 4.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS00234; VWFA; 1.
DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW Signal; Magnesium; Calcium; Repeat.
FT SIGNAL 1 19
FT CHAIN 20 1163 INTEGRIN ALPHA-X.
FT DOMAIN 20 1107 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1108 1128 POTENTIAL.
FT DOMAIN 1129 1163 CYTOPLASMIC (POTENTIAL).
FT REPEAT 34 87 FG-GAP 1.
FT REPEAT ? 351 FG-GAP 2.
FT DOMAIN 165 351 VWFA.
FT REPEAT ? 402 453 FG-GAP 3.
FT REPEAT 455 517 FG-GAP 4.
FT REPEAT 518 576 FG-GAP 5.
FT REPEAT 581 633 FG-GAP 6.
FT REPEAT 466 474 FG-GAP 7.
FT CA_BIND 530 538 POTENTIAL.
FT CA_BIND 593 601 POTENTIAL.
FT SITE 1131 1135 GFRK MOTIF.
FT DISULFID 69 76 BY SIMILARITY.
FT DISULFID 108 126 BY SIMILARITY.
FT DISULFID 655 712 BY SIMILARITY.
FT DISULFID 771 777 BY SIMILARITY.
FT DISULFID 848 863 BY SIMILARITY.
FT DISULFID 998 1022 BY SIMILARITY.
FT DISULFID 1027 1032 BY SIMILARITY.
FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 697 697 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 735 735 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 939 939 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1050 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 490 490 G -> A (IN REF. 2).
FT CONFLICT 756 756 L -> D (IN REF. 2).
FT SEQUENCE 1163 AA; 127885 MW; 6C4E19CC3F62A473 CRC64;

Query Match 9.1%; Score 103; DB 1; Length 1163;
Best Local Similarity 24.5%; Pred. No. 0.39;
Matches 45; Conservative 33; Mismatches 72; Indels 34; Gaps 9;

Qy 2 SFVPSRGTTLMKLTEDRE-----QTRQGLEELQKVLPGGDTVMHEGFERASQIYYEN 56
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 189 SLMOFSNKFQTHFTFEFRRTSNPLSLASVHQLQ-----GFTYTATAIYNVHRLPHAS 243
| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy 57 RQGYRTAS-VIIALTDLGELHEDLFYSREANRSRLGAIVYCVGV-----KDFN 105
| : | : | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 244 YGARDATKILIVTDGKKGDSLDYKD-VIPMDAAGIIRYAIYGVGLAFQNRNSWELN 302
| : | : | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy 106 ETQLARIAD--SKDHVPVNDQFALQGIITHSILKSCIEILAAEPSTICAGESFOVVVR 163
| : | : | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 303 D-----IASKSQEHIKFKVD-FDALKDIQNLKEK-----IFAIEGTETTSSSFELEMA 352
| : | : | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy 164 GNFG 167
| |
Db 353 QEGF 356
```

	Query Match	Best Local Similarity	29.78%; Pred. No. 0.51;	Mismatches	41; Conservative	Indels	23; Gaps
QY	39	TYMHEGFASBQIY-----YENRQGXYTAS--VVIATDGEUHLHEDLFFYSEREANRSRD	91	:	: :	:	:
Dd	346	SYMEKGTMTGALKYLIDNSTVSSEGGAPGAOKGVFTDGSRSD----YINDAKKAKD	401	:	:	:	:
QY	92	LGAIVCYCVGKDVFNTQLARIADS--KDHVPFVNDGFQALOGIIHTSLKKSCIEILLAAEP	149	:	:	:	:
Dd	402	LGFKFMAVGCGNAVEDELREITASEPVAEHYEYTAD-FKTINQIGKKLOKKICVE---EDP	457	:	:	:	:
QY	150	STICAGES---PQVVVRG	164	:	:	:	:
Dd	458	--CACESLVKFOAQKEG	472	:	:	:	:



```
RESULT 15
CALC_RABIT
ID CALC_RABIT STANDARD; PRT; 639 AA.
AC Q28902;
DT 15-JUL-1998 (Rel. 36, Last Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(XII) chain (Fragment).
GN COL2A1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cornea;
RX MEDLINE=95339912; PubMed=7615021;
RA Zhan Q., Burrows R., Cintron C.;
RT "Localization of type XII collagen in normal and healing rabbit
RT cornea by in situ hybridization.";
RL Exp. Eye Res. 60:551-561(1995).
CC -|- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-
CC CONTAINING FIBRILS, THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE
CC SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE
CC LOCALIZED IN THE PERIFIBRILLAR MATRIX (BY SIMILARITY).
CC -|- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 kDa OF
CC NONTRIPLE-HELICAL SEQUENCES (BY SIMILARITY).
CC -|- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT
CC EACH END (BY SIMILARITY).
CC -|- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY
CC SIMILARITY).
CC -|- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY).
CC -|- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -|- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S78179; AAB34889.2; -.
DR HSSP; P02751; ITTF.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00041; fn3; 5.
DR Pfam; PF00092; vwa; 1.
DR SMART; SM00060; FN3; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS0234; VWFA; 1.
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cell adhesion; Collagen; Glycoprotein.
FT NON_TER 1
FT DOMAIN <1 114 VWFA.
FT DOMAIN 127 216 FIBRONECTIN TYPE-III 1.
FT DOMAIN 217 307 FIBRONECTIN TYPE-III 2.
FT DOMAIN 308 395 FIBRONECTIN TYPE-III 3.
FT DOMAIN 397 487 FIBRONECTIN TYPE-III 4.
FT DOMAIN 488 578 FIBRONECTIN TYPE-III 5.
FT DOMAIN 579 >639 FIBRONECTIN TYPE-III 6.
FT NON_TER 639
SQ SEQUENCE 639 AA; 71032 MW; 162C6A923F551E6C CRC64;
```

Query Match 8.48; Score 95; DB 1; Length 639;  
Best Local Similarity 26.18; Pred. No. 0.93;  
Matches 40; Conservative 29; Mismatches 60; Indels 24; Gaps 9;

```
OY 24 RQGLEELQKVLVLP---GGDTY---MHEGFERASEQIYYENRQGY--RTASVIIALTGDELHED 77
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 RKSLLQAVANLPYKGGNTLTGMALNFIR---QQNFKTQAGMRPRARKIGVLTIDGKSODD 58
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 78 LFFYSEREANRSRDLGAIVYCVGVKDFNETQLARTADSKD--HVFPVNDGFOALQGIHHS 135
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 59 V-----EAPSKKLKDEGVLEFAIGIKNADEVELKMIATPDODTHAYNVAD-FDSLSKIYDD 113
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 136 ILKSKSCIEILA-----AEPSTICAGE-----SFQV 160
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 114 LTINLCNSVKGPGDLEAPSNLIVISERTHRSPRV 146
      | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: August 9, 2002, 10:47:09  
Job time: 832 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:34:21 ; Search time 61.73 Seconds  
(without alignments)  
339.340 Million cell updates/sec

Title: US-09-970-076-10  
Perfect score: 1131  
Sequence: 1 MSFTVFSTRGTTLMKLTEDR.....STSGFKBGNHPCLPARPHT 218

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114	10.1	3124	1 A40020	collagen alpha 1(X
2	112	9.9	1153	2 S00551	leukocyte surface
3	111	9.8	1153	1 RWHU1B	cell surface glyco
4	109.5	9.7	1747	1 A45974	collagen alpha 1(X
5	109	9.6	929	2 I51027	type XII collagen
6	106.5	9.4	760	1 C2WS	classical-compleme
7	105.5	9.3	724	2 A48569	antigen Em100 - Ei
8	103	9.1	1163	1 RWHU1C	cell surface glyco
9	102.5	9.1	1857	2 S31212	collagen alpha 1(X
10	102.5	9.1	1888	2 S78476	collagen alpha 1(X
11	99	8.8	698	2 B96958	dnak protein (heat
12	96.5	8.5	496	2 A37979	cartilage matrix p
13	96.5	8.5	817	2 T50240	kinesin-like prote
14	95	8.4	764	1 BBHU	complement factor
15	93.5	8.3	500	2 S65522	cartilage matrix p
16	92.5	8.2	4307	2 T20721	hypothetical prote
17	90	8.0	712	2 A45638	immunodominant mic
18	89	7.9	537	2 T04822	hypothetical prote
19	89	7.9	843	2 A40970	undulin 1 - human
20	88.5	7.8	487	1 V2EBPT	sensor kinase phoQ
21	88	7.8	2944	2 A54849	collagen alpha 1(V
22	87	7.7	689	2 F84811	probable retroelem
23	86.5	7.6	487	2 AG0646	sensor protein Pho
24	85.5	7.6	453	2 H96798	hypothetical prote
25	85.5	7.6	3176	1 CGHU3A	collagen alpha 3(V
26	85	7.5	3051	2 S42373	hypothetical prote
27	84.5	7.5	237	2 E69838	transcription regu
28	84.5	7.5	240	2 G71227	hypothetical prote
29	83.5	7.4	432	1 E70373	conserved hypothet

## ALIGNMENTS

## RESULT 1

A40020  
collagen alpha 1(XII) chain precursor - chicken  
N:Alternate names: fibrochimerin  
C:Species: Gallus gallus (chicken)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 19-Jan-2001  
C:Accession: A40020; A34485; B34485; A28037; S23814; S2254; S28811  
R:Yamagata, M.; Yamada, K.M.; Yamada, S.S.; Shinomura, T.; Tanaka, H.; Nishida, Y.; O  
J. Cell Biol. 115, 209-221, 1991  
A:Title: The complete primary structure of type XII collagen shows a chimeric molecu  
nous region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp si  
A:Reference number: A40020; MUID:92011862  
A:Accession: A40020  
A:Molecule type: mRNA  
A:Residues: 1-3124 <YAM>  
A:Cross-references: GB:D00824; NID:q222810; PIDN:BAA00701.1; PID:q222811  
A:Note: In the authors' translation residues 1216-1219 are shown after residue 1235 a  
R:Gordon, M.K.; Gerecke, D.R.; Dublet, B.; van der Rest, M.; Olsen, B.R.  
J. Biol. Chem. 264, 19772-19778, 1989  
A:Title: Type XII collagen. A large multidomain molecule with partial homology to typ  
A:Reference number: A34485; MUID:90062079  
A:Accession: A34485  
A:Molecule type: mRNA  
A:Residues: 2456-2758 'A', 2760-2802 'F', 2804-2976 'F', 2978-3124 <GOR>  
A:Cross-references: EMBL:J05137; NID:q211284; PIDN:AAA48635.1; PID:q211285  
A:Accession: B34485  
A:Molecule type: protein  
A:Residues: 2772-2792; 2846-2873 <GOR2>  
R:Gordon, M.K.; Gerecke, D.R.; Olsen, B.R.,  
Proc. Natl. Acad. Sci. U.S.A. 84, 6040-6044, 1987  
A:Title: Type XII collagen: distinct extracellular matrix component discovered by CDN  
A:Reference number: A28037; MUID:87317590  
A:Accession: A28037  
A:Molecule type: mRNA  
A:Residues: 2960-2976 'F', 2978-3074 'AG' <GOR>  
A:Cross-references: EMBL:M17375; NID:q211649; PIDN:AAA48718.1; PID:q211650  
A:Note: this sequence has been revised in reference A34485  
R:Koch, M.; Bernasconi, C.; Chiquet, M.  
Eur. J. Biochem. 207, 847-856, 1992  
A:Title: A major oligomeric fibroblast proteoglycan identified as a novel large form  
A:Reference number: S23814; MUID:92362621  
A:Accession: S23814  
A:Molecule type: protein  
A:Residues: 'X', 1333, 'Q', 1335-1347; 1914-1928; 2504, 'X', 2506, 'X', 2508-2511, 'X', 2513-251  
R:Dublet, B.; van der Rest, M.  
J. Biol. Chem. 262, 17724-17727, 1987  
A:Title: Type XII collagen is expressed in embryonic chick tendons. Isolation of pep  
A:Reference number: S2254; MUID:88087065  
A:Accession: S2254  
A:Molecule type: protein  
A:Residues: 2831-2832, 'T', 2834, 'R', 2836-2843; 3002-3014 <DUB>  
R:Trueb, J.; Trueb, B.

hypothetical prote  
hypothetical prote  
transcription regu  
conserved hypothet  
hypothetical prote  
hypothetical prote  
complement factor  
hypothetical prote  
hypothetical prote  
penicillin-binding  
hypothetical prote  
penicillin-binding  
penicillin-binding  
penicillin-binding  
DNA polymerase I -  
spore coat peptide

C:Species: Mus musculus (house mouse)  
C:Date: 30-Sep-1989 #sequence\_revision 30-Sep-1991 #text\_change 22-Oct-1999  
C:Accession: S00551; 159078  
R:Pytela, R.  
EMBO J. 7, 1371-1378, 1988  
A:Title: Amino acid sequence of the murine Mac-1 alpha chain reveals homology with th  
A:Reference number: S00551; MUID:88312584  
A:Accession: S00551  
A:Molecule type: DNA  
A:Residues: 1-1153 <P>  
A:Cross-references: EMBL:X07640; NID:g52982; PIDN:CAA30479.1; PID:g52983  
A:Note: the authors translated the codon CAC for residue 569 as Gln  
R:Sastre, L.; Roman, J.M.; Teplow, D.B.; Dreyer, W.J.; Gee, C.E.; Larson, R.S.; Rober  
proc. Natl. Acad. Sci. U.S.A. 83, 5644-5648, 1986  
A:Title: A partial genomic DNA clone for the alpha subunit of the mouse complement re  
A:Reference number: 159078; MUID:86287312  
A:Accession: 159078  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 11-44 <RES>  
A:Cross-references: GB:M14293; NID:gl98993; PIDN:AAA39484.1; PID:g554193  
C:Genetics:  
A:Gene: Mac-1  
C:Superfamily: cell surface glycoprotein CD1b; von Willebrand factor type A repeat h  
C:Keywords: cell adhesion; glycoprotein; transmembrane protein  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-1153/Product: leukocyte surface glycoprotein Mac-1 alpha chain #status experimen  
F:148-318/Domain: von Willebrand factor type A repeat homology <VWA2>  
F:1106-1129/Domain: transmembrane #status predicted <TM>

Query Match 9.98; Score 112; DB 2; Length 1153;  
Best Local Similarity 29.18; Pred. No. 0.14;  
Matches 41; Conservative 26; Mismatches 56; Indels 18; Gaps 7;  
QY 37 GDTYMHEGFERASQIYYE-NRQGYRTASVIATDGLHEDLFFYSF--REANRSRLG 93  
Db 223 GRTKTASGLRVRFLPHKTNARENAKILVITDGEKFGDPLDYKDVIPEADRA---G 279  
QY 94 AIYVCVGKVD-FNEQTALRIAD-----SKDHVFPVNDGFOALQGIHSLKSCIFELA 146  
Db 280 VIRYVIGVGNFAPKPSRRELTDTASKPAGEHFQV-DNFEALNTIQLOEK----IFA 334  
QY 147 APPSTICAGESFQVVVRNGNF 167  
Db 335 IEGTGTGSTSSFEHEMSQEGF 355  
RESULT 3  
RWHLB  
cell surface glycoprotein CD1b precursor [validated] - human  
N:Alternate names: complement receptor type 3 alpha chain; leukocyte adhesion protein  
eukocyte integrin alpha chain; neutrophil adherence receptor alphaM chain  
C:Species: Homo sapiens (man)  
C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 31-Dec-2000  
C:Accession: A31108; A28915; A41600; A30892; A32218; A46526; A26091; I52567  
R:Corbi, A.L.; Kishimoto, T.K.; Miller, L.J.; Springer, T.A.  
J. Biol. Chem. 263, 12403-12411, 1988  
A:Title: The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3,  
B  
A:Reference number: A31108; MUID:88315033  
A:Accession: A31108  
A:Molecule type: mRNA  
A:Residues: 1-1153 <COR>  
A:Cross-references: GB:J03925; NID:gl87284; PIDN:AAA59544.1; PID:g307148  
A:Note: part of this sequence was confirmed by protein sequencing  
R:Arnaut, M.A.; Gupta, S.K.; Pierce, M.W.; Tenen, D.G.  
J. Cell Biol. 106, 2153-2158, 1988  
A:Title: Amino acid sequence of the alpha subunit of human leukocyte adhesion recepto  
A:Reference number: A28915; MUID:88257215  
A:Accession: A28915  
A:Molecule type: mRNA  
A:Residues: 1-499,501-965,'p',967-1153 <ARN>

Biochim. Biophys. Acta 1171, 97-98, 1992  
A:Title: The two splice variants of collagen XII share a common 5' end.  
A:Reference number: S28811; MUID:93042014  
A:Accession: S28811  
A:Molecule type: mRNA  
A:Status: preliminary  
A:Residues: 1-24,1189-1257,'S',1259-1263,'E',1265-1280 <TRU>  
A:Cross-references: EMBL:X67327  
C:Genetics:  
A:Introns: 2845/3; 2863/3; 2887/3; 2899/3; 2922/1; 2985/1; 3008/1; 3065/1  
C:Superfamily: collagen alpha 1(XII) chain; fibronectin type III repeat homology; von W  
C:Keywords: alternative splicing; cell binding; coiled coil; connective tissue; disulfid  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-3124/Product: collagen alpha 1(XII) chain #status predicted <MAT>  
F:24-1189-3124/Product: collagen alpha 1(XII) chain short splice form #status predicted  
F:24-114/Domain: fibronectin type III repeat homology <FN3A>  
F:137-301/Domain: von Willebrand factor type A repeat homology <VWA1>  
F:332-425/Domain: fibronectin type III repeat homology <FN3B>  
F:332-414/Domain: von Willebrand factor type A repeat homology <VWA2>  
F:437-601/Domain: fibronectin type III repeat homology <FN3C>  
F:629-1178/Domain: fibronectin type III repeat homology <FN3D>  
F:630-711/Domain: fibronectin type III repeat homology <FN3E>  
F:721-802/Domain: fibronectin type III repeat homology <FN3F>  
F:812-895/Domain: fibronectin type III repeat homology <FN3G>  
F:905-1076/Domain: fibronectin type III repeat homology <FN3H>  
F:1086-1169/Domain: fibronectin type III repeat homology <FN3I>  
F:1197-1361/Domain: von Willebrand factor type A repeat homology <VWA3>  
F:1384-2295/Domain: fibronectin type III repeat homology <FN3J>  
F:1474-1557/Domain: fibronectin type III repeat homology <FN3K>  
F:1566-1647/Domain: fibronectin type III repeat homology <FN3L>  
F:1655-1738/Domain: fibronectin type III repeat homology <FN3M>  
F:1756-1838/Domain: fibronectin type III repeat homology <FN3N>  
F:1847-1928/Domain: fibronectin type III repeat homology <FN3O>  
F:1937-2019/Domain: fibronectin type III repeat homology <FN3P>  
F:2028-2110/Domain: fibronectin type III repeat homology <FN3Q>  
F:2119-2199/Domain: fibronectin type III repeat homology <FN3R>  
F:2207-2294/Domain: fibronectin type III repeat homology <FN3S>  
F:2325-2490/Domain: von Willebrand factor type A repeat homology <VWA4>  
F:2438-2440/Region: cell adhesion #status predicted  
F:2509-2750/Domain: collagenous NC2 #status predicted <COL2>  
F:2751-2902/Domain: collagenous NC2 #status predicted <COL1>  
F:2903-2945/Domain: collagenous NC1 #status predicted <COL1>  
F:2946-3048/Domain: non-collagenous NC1 #status predicted <NC1>  
F:3049-3124/Domain: non-collagenous NC1 #status predicted <NC1>  
F:32,1006,1032,1044,1512,1767,2210,2273,2532,2683/Binding site: carbohydrate (Asn) #stat  
F:2780,2789,2836,2842,2860,2866,2869,3004,3007/Modified site: hydroxyproline (Pro) #stat

Query Match 10.18; Score 114; DB 1; Length 3124;  
Best Local Similarity 27.88; Pred. No. 0.34;  
Matches 45; Conservative 27; Mismatches 66; Indels 24; Gaps 8;  
QY 36 GDTYMHEGFERASQIYYENRQGYR--TASVIATDGLHEDLFFYSEREANRSRLG 93  
Db 513 GSTNTGKATYVREKVFVTSK-GSRPNVPRVMILITDCK-SSDAF---KEPAIKLRDAD 567  
QY 94 AIYVCVGKVD-FNEQTALRIAD-----SKDHVFPVNDGFOALQGIHSLKSCIFELAAEPST 151  
Db 568 VEIFAVGVKDAVRTELEATASPAETHVTVVED-FDAQRTISFELTQSVCLRI----- 619  
QY 152 ICAGESFQVVVRNGNFRHARNV---DRVLCSEKINDSVTLK 190  
Db 620 -----EQELAAIRKKSIVPAKNVFSVDTSSEKVSWSAAGSE 657  
RESULT 2  
S00551  
leukocyte surface glycoprotein Mac-1 alpha chain precursor - mouse  
N:Alternate names: complement-3 receptor alpha chain

A:Cross-references: GB:M18044; GB:J03270; GB:M19664; GB:X07421; NID:gl186935; PIDN:AAA594  
A:Note: the authors translated the codon TAC for residue 1129 as Thr  
A:Note: part of this sequence, including the amino end of the mature protein, was confir  
R:Shelley, C.S.; Arnaut, M.A.  
Proc. Natl. Acad. Sci. U.S.A. 88, 10525-10529, 1991  
A:Title: The promoter of the CD11b gene directs myeloid-specific and developmentally reg  
A:Reference number: A41600; MUID:92073318  
A:Accession: A41600  
A:Molecule type: DNA  
A:Residues: 1-9 <SH>  
A:Cross-references: GB:M76724; NID:gl180018; PIDN:AAA58410.1; PID:g553215  
R:Arnaut, M.A.; Remold-O'Donnell, E.; Pierce, M.W.; Harris, P.; Tenen, D.G.  
Proc. Natl. Acad. Sci. U.S.A. 85, 2776-2780, 1988  
A:Title: Molecular cloning of the alpha-subunit of human and guinea pig leukocyte adhesi  
A:Reference number: A94193; MUID:88190151  
A:Accession: A30892  
A:Molecule type: mRNA  
A:Residues: 917-1042 <AR2>  
A:Cross-references: GB:M18044  
R:Hickstein, D.D.; Hickey, M.J.; Ozols, J.; Baker, D.M.; Back, A.L.; Roth, G.J.  
Proc. Natl. Acad. Sci. U.S.A. 86, 257-261, 1989  
A:Title: cDNA sequence for the alpha subunit of the human neutrophil adherence receptor  
A:Reference number: A32218; MUID:89098893  
A:Accession: A32218  
A:Molecule type: mRNA  
A:Residues: 9-1153 <HC>  
A:Cross-references: GB:J04145; NID:gl189068; PIDN:AAA59903.1; PID:g386975  
A:Note: part of this sequence was confirmed by protein sequencing  
R:Fleming, J.C.; Pahl, H.L.; Gonzalez, D.A.; Smith, T.F.; Tenen, D.G.  
J. Immunol. 150, 480-490, 1993  
A:Title: Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-in  
n during evolution.  
A:Reference number: A46526; MUID:93123748  
A:Accession: A46526  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-499,501-1153 <FILE>  
A:Cross-references: GB:S52227; NID:g263047; PIDN:AAB24821.1; PID:g263049  
A:Note: the last three bases of Intron 13, CAG, are included in some but not all mature  
A:Note: sequence extracted from NCBI backbone (NCBIP:121963)  
R:Pierce, M.W.; Remold-O'Donnell, E.; Todd III, R.F.; Arnaut, M.A.  
Biochim. Biophys. Acta 874, 368-371, 1986  
A:Title: N-terminal sequence of human leukocyte glycoprotein Mol: conservation across sp  
A:Reference number: A90664; MUID:87076671  
A:Accession: A26091  
A:Molecule type: protein  
A:Residues: 17-31 <PI>  
A:Experimental source: granulocytes  
R:Pahl, H.L.; Rosmarin, A.G.; Tenen, D.G.  
Blood 79, 865-870, 1992  
A:Title: Characterization of the myeloid-specific CD11b promoter.  
A:Reference number: I52567; MUID:92144986  
A:Accession: I52567  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-9 <RS>  
A:Cross-references: GB:M84477; NID:gl180184; PIDN:AAA51960.1; PID:g553219  
C:Comment: A common beta chain (CD18) forms a heterodimer with this chain to form Mac-1  
C:Genetics:  
A:Gene: GDB:ITGAM; CR3A  
A:Cross-references: GDB:I20599; OMIM:120980  
A:Map position: 16p11.2-16p11.2  
A:Note: promoter contains a GATA motif and two Spl consensus binding sites  
C:Superfamily: cell surface glycoprotein Cn1b; von Willebrand factor type A repeat hom  
C:Keywords: alternative splicing; calcium; cell adhesion; glycoprotein; heterodimer; mag  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-1153/Product: cell surface glycoprotein Cn1b #status experimental <MAT>  
F:17-1108/Domain: extracellular #status predicted <EXT>  
F:148-318/Domain: von Willebrand factor type A repeat homology <VWA2>  
F:465-473/Region: calcium/magnesium binding #status predicted  
F:530-538/Region: calcium/magnesium binding #status predicted  
F:593-601/Region: calcium/magnesium binding #status predicted  
F:1109-1134/Domain: transmembrane #status predicted <TM>

```

F:1135-1153/Domain: intracellular #status predicted <INT>
F:86,240,391,469,693,697,735,802,881,901,912,941,947,979,994,1022,1045,1051,1076/Bind
Query Match          9.8%; Score 111; DB 1; Length 1153;
Best Local Similarity 28.5%; Pred. No. 0.18;
Matches 41; Conservative 27; Mismatches 58; Indels 18; Gaps 7;

QY 37 GDTVMHGEFRASQIYYENRQGYRTA-SVIIATDGHEDHEDLFFYSE--REANRSRLDG 93
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 223 GRHTATGIRKVVRELFENITNGARKNAFKILVVITDGEKFDPLGYEDVPEADRE---G 279

QY 94 AIYVGVGVKDFNETQLAR-----IADS--KDHVPFVNDGFOALQGIITHSLKSCIEILA 146
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 280 VIRYVIGVGDAFRSEKSRQELNTASKPRDHPVQVNN-TEALKTIQNLREK----IFA 334

QY 147 AEPSTICAGESFQVVVRGNGFRHA 170
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 335 IEGTGTGSSSSFEHMSQEGFSAA 358

RESULT 4
A45974
collagen alpha 1(XIV) chain precursor, short form 2 - chicken
N:Alternate names: undulin
C:Species: Gallus gallus (chicken)
C:Date: 10-Sep-1999 #sequence.revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A45974; S30085; S22916; S17035; S20833
R:Garecke, D.R.; Foley, J.W.; Castagnola, P.; Gennari, M.; Dublet, B.; Cancedda, R.;
J. Biol. Chem. 268, 12177-12184, 1993
A:Title: Type XIV collagen is encoded by alternative transcripts with distinct 5' reg-
ns.
A:Reference number: A45974; MUID:93280195
A:Accession: A45974
A>Status: preliminary
A:Molecule type: mRNA; protein
A:Residues: 1-1747 <GER>
A:Experimental source: embryo skin
A>Note: sequence inconsistent with the nucleotide translation
A>Note: sequence extracted from NCBI backbone (NCBIN:133364, NCBI:133365)
R:Apte, S.S.
submitted to the EMBL Data Library, March 1992
A:Reference number: S30085
A:Accession: S30085
A:Molecule type: mRNA
A:Residues: 1472-1660 <APT>
A:Cross-references: EMBL:X55122; NID:G62871; PIDN:CAA46238.1; PID:9338175
R:Trueb, J.; Trueb, B.
Eur. J. Biochem. 207, 549-557, 1992
A:Title: Type XIV collagen is a variant of undulin.
A:Reference number: S22916; MUID:92339443
A:Accession: S22916
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 286-494, 'Q',496-834,'A',836-1119,'KL',1122-1402,1409-1439 <TRU>
R:Gordon, M.K.; Castagnola, P.; Dublet, B.; Linsemayer, T.F.; van der Rest, M.; Mayn-
Eur. J. Biochem. 201, 333-338, 1991
A:Title: Cloning of a cDNA for a new member of the class of fibril-associated collages
A:Reference number: S17035; MUID:92037585
A:Accession: S17035
A:Molecule type: mRNA
A:Residues: 1472-1659 <GOR1>
A:Accession: S20833
A:Molecule type: protein
A:Residues: 1551-1570;1593-1599;1639-1667 <GOR2>
C:Superfamily: collagen alpha 1(XIV) chain; fibronectin type III repeat homology; vonn
C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; tr
F:40-204/Domain: von Willebrand factor type A repeat homology <VWA1>
F:236-317/Domain: fibronectin type III repeat homology <FN3A>
F:326-409/Domain: fibronectin type III repeat homology <FN3B>
F:418-498/Domain: fibronectin type III repeat homology <FN3C>
F:507-591/Domain: fibronectin type III repeat homology <FN3D>
F:625-707/Domain: fibronectin type III repeat homology <FN3E>

```

A:Cross-references: GB:M18044; GB:J03270; GB:M19664; GB:X07421; NID:gl186935; PIDN:AAA594  
A:Note: the authors translated the codon TAC for residue 1129 as Thr  
A:Note: part of this sequence, including the amino end of the mature protein, was confir  
R:Shelley, C.S.; Arnaut, M.A.  
Proc. Natl. Acad. Sci. U.S.A. 88, 10525-10529, 1991  
A:Title: The promoter of the CD11b gene directs myeloid-specific and developmentally reg  
A:Reference number: A41600; MUID:92073318  
A:Accession: A41600  
A:Molecule type: DNA  
A:Residues: 1-9 <SH>  
A:Cross-references: GB:M76724; NID:gl180018; PIDN:AAA58410.1; PID:g553215  
R:Arnaut, M.A.; Remold-O'Donnell, E.; Pierce, M.W.; Harris, P.; Tenen, D.G.  
Proc. Natl. Acad. Sci. U.S.A. 85, 2776-2780, 1988  
A:Title: Molecular cloning of the alpha-subunit of human and guinea pig leukocyte adhesi  
A:Reference number: A94193; MUID:88190151  
A:Accession: A30892  
A:Molecule type: mRNA  
A:Residues: 917-1042 <AR2>  
A:Cross-references: GB:M18044  
R:Hickstein, D.D.; Hickey, M.J.; Ozols, J.; Baker, D.M.; Back, A.L.; Roth, G.J.  
Proc. Natl. Acad. Sci. U.S.A. 86, 257-261, 1989  
A:Title: cDNA sequence for the alpha subunit of the human neutrophil adherence receptor  
A:Reference number: A32218; MUID:89098893  
A:Accession: A32218  
A:Molecule type: mRNA  
A:Residues: 9-1153 <HC>  
A:Cross-references: GB:J04145; NID:gl189068; PIDN:AAA59903.1; PID:g386975  
A:Note: part of this sequence was confirmed by protein sequencing  
R:Fleming, J.C.; Pahl, H.L.; Gonzalez, D.A.; Smith, T.F.; Tenen, D.G.  
J. Immunol. 150, 480-490, 1993  
A:Title: Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-in  
n during evolution.  
A:Reference number: A46526; MUID:93123748  
A:Accession: A46526  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-499,501-1153 <FILE>  
A:Cross-references: GB:S52227; NID:g263047; PIDN:AAB24821.1; PID:g263049  
A:Note: the last three bases of Intron 13, CAG, are included in some but not all mature  
A:Note: sequence extracted from NCBI backbone (NCBIP:121963)  
R:Pierce, M.W.; Remold-O'Donnell, E.; Todd III, R.F.; Arnaut, M.A.  
Biochim. Biophys. Acta 874, 368-371, 1986  
A:Title: N-terminal sequence of human leukocyte glycoprotein Mol: conservation across sp  
A:Reference number: A90664; MUID:87076671  
A:Accession: A26091  
A:Molecule type: protein  
A:Residues: 17-31 <PI>  
A:Experimental source: granulocytes  
R:Pahl, H.L.; Rosmarin, A.G.; Tenen, D.G.  
Blood 79, 865-870, 1992  
A:Title: Characterization of the myeloid-specific CD11b promoter.  
A:Reference number: I52567; MUID:92144986  
A:Accession: I52567  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-9 <RS>  
A:Cross-references: GB:M84477; NID:gl180184; PIDN:AAA51960.1; PID:g553219  
C:Comment: A common beta chain (CD18) forms a heterodimer with this chain to form Mac-1  
C:Genetics:  
A:Gene: GDB:ITGAM; CR3A  
A:Cross-references: GDB:120599; OMIM:120980  
A:Map position: 16p11.2-16p11.2  
A:Note: promoter contains a GATA motif and two Spl consensus binding sites  
C:Superfamily: cell surface glycoprotein Cc1b; von Willebrand factor type A repeat hom  
C:Keywords: alternative splicing; calcium; cell adhesion; glycoprotein; heterodimer; mag  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-1153/Product: cell surface glycoprotein Cc1b #status experimental <MAT>  
F:17-1108/Domain: extracellular #status predicted <EXT>  
F:148-318/Domain: von Willebrand factor type A repeat homology <VWA2>  
F:465-473/Region: calcium/magnesium binding #status predicted  
F:530-538/Region: calcium/magnesium binding #status predicted  
F:593-601/Region: calcium/magnesium binding #status predicted  
F:1109-1134/Domain: transmembrane #status predicted <TM>



```

Qy      57 RQGYRTAS-VIIALTGDELHEDLFFYSERENANRSDLGAIYVCV-----KDFN 105
         ||| :||| | | | | | | | | | | | | | | | | | | | | | |
Db      244 YGARRDATKILIVITDGKKEGDSDLDKD-VIPMADAAGIIRYAIGVGFLAFQNRRNSMKELN 302
         ||| :||| | | | | | | | | | | | | | | | | | | | | | |
Qy      106 ETQLARIAD--SKDHVPVNDGFOALOGIIHSILKKSCEILAAPETICAGESFQVVVR 163
         :   ||| :||| | | | | | | | | | | | | | | | | | | | | | |
Db      303 D-----TASPSQEHIKFVED-FDALKDIONQLKEK---IFAIEGTETTSSSFELEMA 352
         :   ||| :||| | | | | | | | | | | | | | | | | | | | | | |
Qy      164 GNGF 167
         ||
Db      353 QEGF 356

```

C:Species: Gallus gallus (Chicken)  
C>Date: 11-Mar-1998 #sequence\_revision 17-Apr-1998 #text\_change 13-Aug-1999  
C:Accession: S31212  
R:Waelchli, C.; Trueb, B.; Winterhalter, K.H.; Trueb, B.  
Eur. J. Biochem. 212, 483-490, 1993  
A:Title: Complete primary structure of chicken collagen XIV.  
A:Reference number: S31211; MUID:93185668  
A:Accession: S31212  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-1857 <WAE>  
A:CROSS-references: EMBL:X70792; NID:g288874; PID:nCAA50063.1; PID:g288875  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
C:Genetics:  
A:Gene: Coll14A1  
C:Superfamily: collagen alpha 1(XIV) chain; fibronectin type III repeat homology; von  
C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; tr  
F:1-28/Domain: signal sequence #status predicted <SIG>  
F:28-1857/Product: collagen alpha 1(XIV) chain, short form #status predicted <MAT>  
F:29-110/Domain: fibronectin type III repeat homology <FN3A>  
F:156-320/Domain: von Willebrand factor type A repeat homology <VWA1>  
F:352-433/Domain: fibronectin type III repeat homology <FN3B>  
F:442-525/Domain: fibronectin type III repeat homology <FN3C>  
F:534-614/Domain: fibronectin type III repeat homology <FN3D>  
F:623-707/Domain: fibronectin type III repeat homology <FN3E>  
F:741-823/Domain: fibronectin type III repeat homology <FN3F>  
F:832-914/Domain: fibronectin type III repeat homology <FN3G>  
F:922-1009/Domain: fibronectin type III repeat homology <FN3H>  
F:1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>

QY 115 -SKDHVPVNDGFOALQGLIIHSILKKSCEIILAAEPSILCAEESI QYVVGROGRIIMKART

F;1-22/Domain: signal sequence #status predicted



F:23-496/Product: cartilage matrix protein #status predicted <MAT>  
F:39-206/Domain: von Willebrand factor type A repeat homology <VWAL>  
E:227-262/Domain: EGF homology <EG>  
F:273-437/Domain: von Willebrand factor type A repeat homology <VWAA>  
F:76,344/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:221-238,234-247,249-262/Disulfide bonds: #status predicted

Query Match 8.5%; Score 96.5; DB 2; Length 496;  
Best Local Similarity 29.7%; Pred. No. 1.2;  
Matches 41; Conservative 19; Mismatches 55; Indels 23; Gaps 8;  
QY 39 TYMHGFFERASQIY-----YENROGYRTAS--VIAITDGDHEDLFFYSREANRSRD 91  
DB 346 SYMEKGTWTGAALKYLIIDNSFTVSSGARPGAKVGIVFTDGRSQD----YINDAAKKAKD 401  
QY 92 LGAIYVCVGVKDFNETQLARTADS--KDHVPVNDGFOALQGIHSLKSCIEILAEAP 149  
DB 402 LGFKMFVAVGVGNVAVDELRETASEPVAEHFYFTAD-FKTIINQIGKKLQKKICVE---EDP 457  
QY 150 STICAGES---FOVVVRG 164  
DB 458 ---CACESLVKFOAKVEG 472

RESULT 13  
T50240  
Kinesin-like protein [imported] - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 09-Jun-2000 #sequence revision 09-Jun-2000 #text\_change 09-Jun-2000  
C:Accession: T50240  
R:Cadieu, E.; Lelaure, V.; Galibert, F.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, January 1999  
A:Reference number: Z25048  
A:Accession: T50240  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-817 <CAD>  
A:Cross-references: EMBL:AL136235; PIDN:CAB65811.1; GSPDB:GN00066; SPDB:SPAC664.10  
A:Experimental source: strain 972h(-); cosmid c664  
C:Genetics:  
A:Gene: SPDB:SPAC664.10  
A:Map position: 1  
A:Introns: 10/3

Query Match 8.5%; Score 96.5; DB 2; Length 817;  
Best Local Similarity 25.1%; Pred. No. 2.2;  
Matches 51; Conservative 33; Mismatches 88; Indels 31; Gaps 10;  
QY 11 TFLMKLTEDREIQIGLEQLKVLPGDITYMHGFFERASEQIYYENROGYRTASVIALT 70  
DB 360 SAIQPLOENNSLKQIEQLQREL-ASEYIVYKENLKSSLDQ-QSANVQKLESTNALEST 417  
QY 71 DGHEDLFFYSE-----REANRSRLGAIYVCVGVKDFNETQLARIADSKDH--V 119  
DB 418 IKTLEDVYTMKNTIELGLKLSANVERD-GLVEKLIAEETLRRKLHNTIQELKGNIRV 476  
QY 120 F-----PVNDGFOALQGIHSLKSCIEILAEFSTICAGSFQVVVGRNGFR-HARNV 173  
DB 477 FCRVRPPLDGSQAFTAPDQNSEASTIEVAQAP-----GSS----LTGNGIKQYAFNF 527  
QY 174 DRVLCFSKINDSV--TLSKLSQS 194  
DB 528 DRVFSPTTNDVFNELSOLIOS 550

RESULT 14  
BBHU  
Complement factor B precursor [validated] - human  
N:Alternate names: C3 convertase; C3 proactivator; glycine-rich beta-glycoprotein; heat-  
N:Contains: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) Bb fragment  
C:Species: Homo sapiens (man)

C:Date: 19-Feb-1984 #sequence\_revision 05-Aug-1994 #text\_change 08-Dec-2000  
C:Accession: S34075; A44622; A00934; A19188; A19947; B19947; B25971; S14339; A44628;  
R:Mejia, J.E.; Jahn, I.; de la Salle, H.; Hauptmann, G.  
submitted to the EMBL Data Library, March 1993  
A:Reference number: S34075  
A:Accession: S34075  
A:Molecule type: mRNA  
A:Residues: 1-764 <MEJ>  
A:Cross-references: EMBL:X72875; NID:G2975568; PIDN:CAA51389.1; PID:G2975569  
R:Woods, D.E.; Markham, A.F.; Ricker, A.T.; Goldberger, G.; Colten, H.R.  
Proc. Natl. Acad. Sci. U.S.A. 79, 5661-5665, 1982  
A:Title: Isolation of cDNA clones for the human complement protein factor B, a class  
A:Reference number: A44622; MUID:83039428  
A:Accession: A44622  
A:Molecule type: mRNA  
A:Residues: 467-546;550-595;752-764 <WOO>  
A:Cross-references: GB:J00185; GB:J00186  
A:Note: the authors translated the codon TAC at 519 as Thr; the nucleic acid translat  
R:Mole, J.E.; Anderson, J.K.; Davison, E.A.; Woods, D.E.  
J. Biol. Chem. 259, 3407-3412, 1984  
A:Title: Complete primary structure for the zymogen of human complement factor B.  
A:Reference number: A20751; MUID:84161997  
A:Accession: A00934  
A:Molecule type: protein; mRNA  
A:Residues: 26-764 <MOL>  
A:Cross-references: GB:K01566  
A:Note: nucleic acid translation differs from the sequence shown in having 300-Leu, 3  
A:Note: 736-Ser was also found  
A:Note: glycosylation sites were determined  
R:Christie, D.L.; Gagnon, J.  
Biochem. J. 209, 61-70, 1983  
A:Title: Amino acid sequence of the Bb fragment from complement factor B. Sequence of  
A:Reference number: A19188; MUID:83204002  
A:Contents: the final paper in a series documenting the sequence, glycosylation site,  
A:Accession: A19188  
A:Molecule type: protein  
A:Residues: 260-296,'T',298-764 <CHR>  
R:Campbell, R.D.; Porter, R.R.  
Proc. Natl. Acad. Sci. U.S.A. 80, 4464-4468, 1983  
A:Title: Molecular cloning and characterization of the gene coding for human compleme  
A:Reference number: A19947; MUID:83273641  
A:Accession: A19947  
A:Molecule type: DNA  
A:Residues: 346-764 <CAM>  
A:Cross-references: GB:J00125  
A:Accession: B19947  
A:Molecule type: mRNA  
A:Residues: 339-509 <CAL>  
A:Cross-references: GB:J00126; NID:g187723; PIDN:AAA36226.1; PID:g553536  
R:Wu, L.; Morley, B.J.; Campbell, R.D.  
Cell 48, 331-342, 1987  
A:Title: Cell-specific expression of the human complement protein factor B gene: evid  
A:Reference number: A25971; MUID:87102880  
A:Accession: B25971  
A:Molecule type: DNA  
A:Residues: 1-99 <MUL>  
A:Cross-references: GB:M15082; NID:g187699; PIDN:AAA59625.1; PID:g553534  
R:Niemann, M.A.; Brown, A.S.; Miller, E.J.  
Biochem. J. 274, 473-480, 1991  
A:Title: The principal site of glycation of human complement Factor B.  
A:Reference number: S14339; MUID:91174758  
A:Accession: S14339  
A:Molecule type: protein  
A:Residues: 270-329 <NIE>  
A:Note: binding site for carbohydrate to lysine under artificial conditions  
R:Morley, B.J.; Campbell, R.D.  
EMBO J. 3, 153-157, 1984  
A:Title: Internal homologies of the Ba fragment from human complement component facto  
A:Reference number: A44628; MUID:84158524  
A:Accession: A44628  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 16-225,'F',227-259 <MOR>

A: Molecule type: cDNA  
A: Residues: 1-764 <RES>  
A: Cross-references: GB:S67310; NID:q452937; PIDN:AA013989.1; PID:q4261689  
R: Horuchi, T.; Kim, S.; Matsumoto, M.; Watanabe, I.; Fujita, S.; Volanakis, J.E.  
Mol. Immunol. 30, 1587-1592, 1993  
A: title: Human Complement factor B: cDNA cloning, nucleotide sequencing, phenotypic con  
A: Reference number: 157824; MUID:94067177  
A: Accession: 157824  
A: Status: translated from GB/EMBL/DBJ  
A: Molecule type: mRNA  
A: Residues: 1-31, 'Q', 33-764 <RES>  
A: Cross-references: GB:L15702; NID:g291921; PIDN:AAA16820.1; PID:g291922  
C: Comment: 292-Cvs has a free sulfhydryl.

A:Note: gene is located in the major histocompatibility complex, class III region  
A:Note: complement factor B initially forms an inactive complex with complement factor C3b forming active C3/C5 convertase; Ba is released  
C:Function:  
A:Description: Bb is a serine proteinase; C3/C5 convertase cleaves complement C3 alpha chain  
A:Pathway: complement alternate pathway  
C:Superfamily: complement C2; complement factor H repeat homology; trypsin homology; von Willebrand factor type A repeat homology; glycoprotein; hydrophobic motif  
C:Keywords: acute phase; complement alternate pathway; duplication;  
F.1-25/Domain: signal sequence #status predicted <SIG>  
F.26-764/Product: complement factor B #status experimental <MAT>  
F.26-259/Product: complement factor Ba fragment #status experimental <BAF>  
F.37-98/Domain: complement factor H repeat homology <PH1>  
F.103-158/Domain: complement factor H repeat homology <PH2>  
F.165-218/Domain: complement factor H repeat homology <PH3>  
F.260-764/Product: C3/C5 convertase Bb fragment #status experimental <BBF>  
F.268-458/Domain: von Willebrand factor type A repeat homology <VFA>  
F.488-752/Domain: trypsin homology #status atypical <TRI>  
F.37-76,62-98,103-145,131-158,165-205,191-218,478-536,511-527,599-615,656-682,695-725/D  
F.122,142,285,378/Binding site: carboxydrate (Asn) (covalent) #status experimental  
F.259-260/Cleavage site: Arg-Lys (complement factor D) #status experimental  
F.526-576,608/Active site: His, Asp, Ser #status experimental

[illegible]



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:32:14 ; Search time 119.13 Seconds  
(without alignments)  
203.258 Million cell updates/sec

Title: US-09-970-076-10  
Perfect score: 1131  
Sequence: 1 MSFIVSTRGTTLMKLTEDR.....STSGFKGNHPCLPARPH 218

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11107396 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL1/gcgdata/hold-geneseg/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1131	100.0	218	22 AAB92985	Human protein sequ
2	1131	100.0	297	22 AAM38976	Human Polypeptide
3	966	85.4	333	21 AAB01422	Human TANGO 197.
4	966	85.4	403	22 AAE01439	Human gene 4 encod
5	961	85.0	403	22 AAE01469	Human gene 4 encod
6	506	44.7	488	21 AAB18447	Amino acid sequenc
7	506	44.7	488	21 AAB18455	A human TANGO 216
8	506	44.7	488	21 AAB18456	A human TANGO 216
9	506	44.7	488	21 AAB18457	A human TANGO 216
10	505	44.7	587	22 AAU19662	Human novel extrac
11	495	43.8	487	21 AAB18448	Amino acid sequenc

12	495	43.8	487	21	AAB18458	A murine TANGO 216
13	495	43.8	487	21	AAB18459	A murine TANGO 216
14	495	43.8	487	21	AAB18460	A murine TANGO 216
15	415	36.7	381	21	AAB01428	Murine TANGO 197.
16	128.5	11.4	1155	16	AAR78167	Mouse alpha-d subu
17	128.5	11.4	1155	18	AAW23060	Mouse beta 2 integ
18	128.5	11.4	1155	19	AAW72835	Mouse alpha-d #1.
19	128.5	11.4	1155	19	AAW65102	Mouse beta-integri
20	128.5	11.4	1155	19	AAW60002	Mouse alpha d poly
21	128.5	11.4	1155	20	AAW73346	Mouse alphad prote
22	128.5	11.4	1155	21	AAW07372	Mouse alpha_d prot
23	128.5	11.4	1161	16	AAR78168	Mouse alpha-d subu
24	128.5	11.4	1161	18	AAW23061	Mouse beta 2 integ
25	128.5	11.4	1161	19	AAW72836	Mouse alpha-d #2.
26	128.5	11.4	1161	19	AAW65103	Mouse beta-integri
27	128.5	11.4	1161	19	AAW60003	Mouse alpha d poly
28	128.5	11.4	1161	20	AAW73347	Mouse alphad prote
29	128.5	11.4	1161	21	AAW07373	Mouse alpha_d prot
30	122	10.8	1151	16	AAR78179	Rat alpha-d compos
31	122	10.8	1151	18	AAW23059	Rat beta 2 integri
32	122	10.8	1151	19	AAW72834	Rat alpha-d #2. R
33	122	10.8	1151	19	AAW65101	Rat beta-integrin
34	122	10.8	1151	19	AAW60001	Rat alpha d polype
35	122	10.8	1151	20	AAW73344	Rat alphad protein
36	122	10.8	1151	21	AAW07371	Rat alpha_d partia
37	122	10.8	1161	16	AAR78169	Rat alpha-d subuni
38	122	10.8	1161	18	AAW23062	Rat beta 2 integri
39	122	10.8	1161	19	AAW72824	Rat alpha-d #1. R
40	122	10.8	1161	19	AAW65104	Rat beta-integrin
41	122	10.8	1161	19	AAW60004	Rat alpha d polype
42	122	10.8	1161	20	AAW73345	Rat alphad protein
43	122	10.8	1161	21	AAW07374	Rat alpha_d protei
44	114	10.1	1161	16	AAR78166	Human beta-2 integ
45	114	10.1	1161	18	AAW23064	Human beta 2 integ

## ALIGNMENTS

RESULT 1  
AAB92985  
ID AAB92985 standard; Protein: 218 AA.  
XX  
AC AAB92985;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:11706.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN EPI074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
(HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX Claim 8; SEQ ID 11706; 2537pp + CD ROM; English.  
PS  
XX The present invention describes primer sets for synthesising 5602  
XX full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC of an oligonucleotide comprising a sequence complementary to the  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are also useful for the  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX Sequence 218 AA;

Query Match 100.0%; Score 1131; DB 22; Length 218;  
Best Local Similarity 100.0%; Pred. No. 3e-124;  
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSFVSTGRTTLMKLTEDREQIROGLEELQKVLPGDGTVMHEGFERASEQIYYENRGY 60  
DB 1 msfivstrgttlmkltedreqirggleelqkvlpgdgtvmhegferaseqiyenrgy 60  
QY 61 RTASVIALTDGELHEDLFFYSERANRDLGAIVYCVGVKDFNETQLARIADSKDHVF 120  
DB 61 rtasvialtdgelhedlffysereanrdrldgaivycvgvkdnetqlariadskdhvf 120  
QY 121 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQVVRGNGFRHARNVDRVLCSF 180  
DB 121 pvndgfoalqgihsllkscieilaaepsticagesfqvvrngfrharnvdrvlcsf 180  
QY 181 KINDSVTLKSLQSPWVSSTSGFKGNSHPCLPARPHT 218  
DB 181 kindsvtlkslqspwvsstsgfkgnshpclparpht 218

RESULT 2  
AAM38976  
ID AAM38976 standard; Protein; 297 AA.  
XX  
XX AAM38976;  
XX  
XX 22-OCT-2001 (first entry)  
XX Human polypeptide SEQ ID NO 2121.  
XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
XX Homo sapiens.  
XX  
XX WO200153312-A1.

XX PD 26-JUL-2001.  
XX PF 26-DEC-2000; 2000WO-US34263.  
XX PR 21-JAN-2000; 2000US-0488725.  
XX PR 25-APR-2000; 2000US-0552317.  
XX PR 09-JUL-2000; 2000US-0598042.  
XX PR 19-JUL-2000; 2000US-0620312.  
XX PR 03-AUG-2000; 2000US-0653450.  
XX PR 14-SEP-2000; 2000US-0662191.  
XX PR 19-OCT-2000; 2000US-0693036.  
XX PR 29-NOV-2000; 2000US-0727344.  
XX (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX WPI: 2001-442253/47.  
XX N-PSDB; AAI58132.  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
XX such as central nervous system injuries -  
XX Example 4; SEQ ID NO 2121; 10078pp; English.  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
XX the encoded polypeptides (AAM38642-AAAM42213) with nontropic,  
XX immunosuppressant and cytostatic activity. The polynucleotides are useful  
XX in gene therapy. A composition containing a polypeptide or polynucleotide  
XX of the invention may be used to treat diseases of the peripheral nervous  
XX system, such as peripheral nervous injuries, peripheral neuropathy and  
XX localised neuropathies and central nervous system diseases, such as  
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
XX utilisation of the activities such as: Immune system suppression,  
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
XX assays for receptor activity, arthritis and inflammation, leukaemias and  
XX C.N.S disorders.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification.  
XX Sequence 297 AA;  
QY  
Query Match 100.0%; Score 1131; DB 22; Length 297;  
Best Local Similarity 100.0%; Pred. No. 5e-124;  
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSFVSTGRTTLMKLTEDREQIROGLEELQKVLPGDGTVMHEGFERASEQIYYENRGY 60  
DB 80 msfivstrgttlmkltedreqirggleelqkvlpgdgtvmhegferaseqiyenrgy 139  
QY 61 RTASVIALTDGELHEDLFFYSERANRDLGAIVYCVGVKDFNETQLARIADSKDHVF 120  
DB 140 rtasvialtdgelhedlffysereanrdrldgaivycvgvkdnetqlariadskdhvf 199  
QY 121 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQVVRGNGFRHARNVDRVLCSF 180  
DB 200 pvndgfoalqgihsllkscieilaaepsticagesfqvvrngfrharnvdrvlcsf 259  
QY 181 KINDSVTLKSLQSPWVSSTSGFKGNSHPCLPARPHT 218  
DB 260 kindsvtlkslqspwvsstsgfkgnshpclparpht 297  
RESULT 3  
AAB01422  
ID AAB01422 standard; Protein; 333 AA.  
XX

AC AAB01422;  
 XX  
 DT 20-OCT-2000 (first entry)  
 XX  
 DE Human TANGO 197.  
 XX  
 KW TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;  
 KW graft versus-host diseases; rheumatoid arthritis; psoriasis;  
 KW inflammatory bowel disease; septic shock; ulcerative colitis;  
 KW Crohn's disease; chronic myelogenous leukemia; cancer; liver  
 KW disease; Hodgkin's disease; osteoarthritis; Lyme's disease;  
 KW cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;  
 KW systemic lupus erythematosus; transgenic animal; diagnosis;  
 KW prognosis; prophylactic; therapeutic; human.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200039284-A1.  
 XX  
 PD 06-JUL-2000.  
 XX  
 PF 23-DEC-1999; 99WO-US31025.  
 XX  
 PR 30-DEC-1998; 98US-0223546.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Holtzman DA;  
 XX  
 DR WPI; 2000-465743/40.  
 DR N-PSDB; AAA47455.  
 XX  
 PT Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,  
 PT 224 and 239 polypeptides useful for the treatment of asthma, rheumatoid  
 PT arthritis, psoriasis and autoimmune diseases  
 XX  
 PS Claim 8; Fig 4; 209pp; English.  
 XX  
 CC Nucleic acids encoding TANGO polypeptides are useful as modulating  
 CC agents for regulating cellular processes like asthma, graft  
 CC versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory  
 CC bowel disease, septic shock, ulcerative colitis, Crohn's disease,  
 CC chronic myelogenous leukemia, cancer, liver disease, Hodgkin's  
 CC disease, osteoarthritis, Lyme's disease, cachexia and autoimmune  
 CC diseases e.g. myasthenia gravis, autoimmune diabetes and systemic  
 CC lupus erythematosus. The nucleic acids are also useful for producing  
 CC transgenic animals and the TANGO polypeptides themselves. Partial  
 CC TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in  
 CC forensic biology, for diagnostic assays, prognostic assays,  
 CC pharmacogenomics and for monitoring clinical trials, TANGO  
 CC polypeptides are suitable for both prophylactic and therapeutic  
 CC methods for treating a subject at risk of a disorder or having a  
 CC disorder associated with aberrant TANGO expression. A wide range  
 CC of cellular disorders can be treated.  
 XX  
 SQ Sequence 333 AA;  
 XX  
 Query Match 85.4%; Score 966; DB 21; Length 333;  
 Best Local Similarity 98.9%; Pred. No. 1.5e-104;  
 Matches: 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSPIVFSTGTTLMKLTDEQIRQGLEELQKVLPGDGYMHGFRASEQIYYENRGY 60  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 80 msfivfstrgttlmkltedrqrlqleelqkvlpgdgytmhgfgraseqiyenrgy 139  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 61 RTASVIALTDGELHEDLFFYSREANRSDLGAIYVCYGVKDFNETQLARIADSKDHFV 120  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 140 rtasvialtdgelhdllffysreanrstdlgaiyvcygvkdfnetqlariadskdhvf 199  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 121 PVNDGFOALQGIHHSILKSCIEILAEFSTCAGESFQVVRNGFRARNVDRVLCSF 180  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 200 pvndgfoalqgiihsilksctieilaefstcagesfqvvrngfrharnvdrvlcsf 259

QY 181 KINDSVTLSK 190  
 DB 260 kindsvtlne 269  
 RESULT 4  
 AAE01439  
 ID AAE01439 standard; Protein; 403 AA.  
 XX  
 AC AAE01439;  
 XX  
 DT 17-JUL-2001 (first entry)  
 XX  
 DE Human gene 4 encoded secreted protein HMLFR02, SEQ ID NO:94.  
 XX  
 KW Human; secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder;  
 KW endocrine disorder; infection; wound healing; vulnery;  
 KW cell culture; chemotaxis; food additive; gene therapy;  
 KW binding partner identification; chromosome 19.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..27  
 FT Protein /label= Signal\_peptide  
 FT 28..403  
 FT /note= "Mature human secreted protein"  
 XX  
 PN WO200134626-A1.  
 XX  
 PD 17-MAY-2001.  
 XX  
 PF 01-NOV-2000; 2000WO-US30045.  
 XX  
 PR 05-NOV-1999; 98US-0163581.  
 PR 30-JUN-2000; 2000US-0215133.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ruben SM, Komatsoulis GA, Moore PA, Birse CE, Ni J;  
 DR WPI; 2001-308778/32.  
 DR N-PSDB; AAD05303.  
 XX  
 PT New nucleic acid molecules encoding 28 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives -  
 XX  
 PS Claim 11; Page 485-486; 562pp; English.  
 XX  
 CC AAD05300-AAD05379 represent cDNAs corresponding to 28 human secreted  
 CC protein genes, and AAE01436-AAE01513 represent the proteins they encode.  
 CC AAE01514-AAE01544 represent human secreted protein fragments or variants.  
 CC The genes and their secreted proteins are useful for preventing,  
 CC treating or ameliorating medical conditions, e.g., by protein or gene  
 CC therapy. Pathological conditions can be diagnosed by determining the  
 CC amount of the new protein in a sample or by determining the presence of  
 CC mutations in the new genes. Specific uses are described for each of the  
 CC 28 genes, based on the tissues in which they are most highly expressed,  
 CC and include developing products for the diagnosis or treatment of  
 CC proliferative disorders, cancer, tumours, foetal and developmental  
 CC abnormalities, haematopoietic disorders, diseases of the immune system,  
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
 CC allergies, neurological disorders (e.g., Alzheimer's disease,  
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,

1 MSFVSTGRGTTLMKLTDEQIRQGLEELQKVLPGDGYMHGEPERASEQIYYENRGY 60  
 |||||  
 80 msfivfstrgttlmkltedreqirggleelqkvlpggdtymhgeraserqiyynrgy 139  
 |||||  
 61 RTASVIALTDGELHEDLFYFSREANRSDLGAIYCVGVKDFNETOLARIADSKDHFV 120  
 |||||  
 QY QY Db QY

FT	Misc-difference	320	Unknown
rm	/label=		

J3



Db 140 rtasvialtdgelhedlffysereansrdlgaixycvgvkdfnetqlariadskdhvf 199  
QY 121 PVNDGFOALQGIHSILKSCIEILAAPSTTCAGSFQVVVRGNGFRHARNVDRVLCSEF 180  
Db 200 pvndgfgalqgihihllkscieilaaepsttcagesfqvvvrngngfrharnvdrvlcsf 259  
QY 181 KINDSVTLSK 190  
Db 260 kindsvtlne 269  
RESULT 6  
ID AAB18447 standard; Protein; 488 AA.  
XX AC AAB18447;  
XX 15-JAN-2001 (first entry)  
XX Amino acid sequence of human TANGO 216 polypeptide.  
XX TANGO 266; TANGO 216; TANGO 261; TANGO 262; TANGO 267;  
KW cellular proliferation; cellular differentiation; cellular adhesion;  
KW von Willebrand factor-associated disorder; cell trafficking; cancer;  
KW hematopoietic associated disease; atelectasis; pulmonary congestion;  
KW oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;  
KW intestinal disorder; spleen associated disease; renal disorder;  
KW cardiovascular disorder; ischemic heart disease; hydrocephalus;  
KW brain herniation; iatrogenic disease; inflammation; meningitis;  
KW Alzheimer's Disease; cerebral toxoplasmosis; Parkinson's disease;  
KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.  
OS Homo sapiens.  
XX FH Key Location/Qualifiers  
FT Peptide 1..34 /note= "signal peptide"  
FT Domain 34..79 /note= "extracellular domain"  
FT Protein 35..488 /note= "mature protein"  
FT Domain 44..213 /note= "von Willebrand factor A domain"  
FT Domain 80..97 /note= "transmembrane domain"  
FT Domain 98..317 /note= "cytoplasmic domain"  
FT Domain 318..341 /note= "transmembrane domain"  
FT Domain 342..488 /note= "extracellular domain"  
XX WO200052022-A1.  
XX 08-SEP-2000.  
XX 01-MAR-2000; 2000WO-US05226.  
XX 01-MAR-1999; 99US-0122458.  
XX (MILL-) MILLENNIUM PHARM INC.  
XX Barnes TM, Holtzman DA, Sharp JD, Fraser CC;  
XX WPI: 2000-579269/54.  
XX N-PSDB; AAA75149.  
XX Novel human and murine secreted proteins designated TANGO 216, 261,  
PT 262, 266 and 267 useful as modulating agents of cellular processes,  
PT e.g. for treating cancer -  
XX Claim 8; Fig 1A-C; 175pp; English.  
XX

CC The present sequence represents a human TANGO 216 polypeptide. The  
CC specification also describes TANGO 266, TANGO 261, TANGO 262, and  
CC TANGO 267. The TANGO polypeptides can be used to modulate cellular  
CC proliferation, modulate cellular differentiation and/or modulate  
CC cellular adhesion. The proteins can be used to treat any von Willebrand  
CC factor-associated disorder, regulate extracellular matrix structuring,  
CC cellular adhesion, and cell trafficking and/or migration, modulate  
CC cellular interactions, modulate cell adhesion in proliferative  
CC disorders, such as cancer, modulate the proliferation, differentiation,  
CC and/or function of cells that appear in the bone marrow, and leukocytes,  
CC treat bone marrow, blood and hematopoietic associated diseases and  
CC disorders, atelectasis, pulmonary congestion or oedema, emphysema,  
CC chronic bronchitis, bronchial asthma and bronchiectasis, intestinal  
CC disorders, spleen associated diseases, modulate renal disorders, treat  
CC cardiovascular disorders such as ischemic heart disease, modulate the  
CC proliferation, differentiation, and/or function of bone and cartilage  
CC cells and to treat bone and/or cartilage associated diseases or  
CC disorder. They may also be used to treat disorders associated with the  
CC ovaries, cerebral oedema, hydrocephalus, brain herniations, iatrogenic  
CC disease, inflammations, bacterial and viral meningitis, Alzheimer's  
CC disease, cerebral toxoplasmosis, Parkinson's disease, multiple sclerosis,  
CC brain cancers, hydrocephalus and encephalitis, and treat hepatic  
CC disorders.  
XX SQ Sequence 488 AA;  
Query Match 44.7%; Score 506; DB 21; Length 488;  
Best Local Similarity 53.4%; Pred. No. 3.3e-50;  
Matches 101; Conservative 37; Mismatches 49; Indels 2; Gaps 1;  
QY 1 MSFVFSVTRGTTLMKLTDEQIRQGLEELQKVLPGDGYMHGFEFASQIYYENRQCY 60  
Db 80 lsfivfssqatllptltdrgkiskgledlkrvspgetyhegiklaneql--qkaggl 137  
QY 61 RTASVIALTDGELHEDLFFYSEREANSRDLGAIVYCVGVKDFNETQLARTADSKDHVF 120  
Db 138 ktssilialtdgldglvpsyaekaisrslgasvycvgvldfefaqleriadskeqvf 197  
QY 121 PVNDGFOALQGIHSILKSCIEILAAPSTTCAGSFQVVVRGNGFRHARNVDRVLCSEF 180  
Db 198 pvkggfgalkglnsilagsctelldqpsvvcvgeefqivisrgfmlgsrngsvlcty 257  
QY 181 KINDSVTLS 189  
Db 258 tvnetytts 266  
RESULT 7  
ID AAB18455 standard; Protein; 488 AA.  
XX AC AAB18455;  
XX 15-JAN-2001 (first entry)  
XX A human TANGO 216 polypeptide clone.  
XX TANGO 266; TANGO 216; TANGO 261; TANGO 262; TANGO 267;  
KW cellular proliferation; cellular differentiation; cellular adhesion;  
KW von Willebrand factor-associated disorder; cell trafficking; cancer;  
KW hematopoietic associated disease; atelectasis; pulmonary congestion;  
KW oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;  
KW intestinal disorder; spleen associated disease; renal disorder;  
KW cardiovascular disorder; ischemic heart disease; hydrocephalus;  
KW brain herniation; iatrogenic disease; inflammation; meningitis;  
KW Alzheimer's Disease; cerebral toxoplasmosis; Parkinson's disease;  
KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.  
XX OS Homo sapiens.  
XX PN WO200052022-A1.  
XX



Db 80 lsfivfssqatllpltdgrgkiskgledlkrvsvpgetyihbegiklaneqi--qkaggl 137  
QY 61 RTASVLIATDCELHEDLFFYSERANRSRDIGAIYCVGVKDFNETQLARTADSKDHVF 120  
Db 138 kssliialtdgldglvpsyaekaisrlgasvycvvgldfefaqleriadskeqvf 197  
QY 121 PVDGFOALQGIHISILKSCIEILAAEPSTICAGESFOVVYRGNGFRHARNVDRVLCSE 180  
Db 198 pvkggfgalkginsilaqstcelldqpssvcvgeefqivlsgrgfmglngsvlcty 257  
QY 181 KINDSVTLS 189  
Db 258 tvnetytts 266

RESULT 9  
AAB18457  
ID AAB18457 standard; Protein; 488 AA.  
XX  
AC AAB18457;  
XX  
DT 15-JAN-2001 (first entry)  
XX  
DE A human TANGO 216 polypeptide clone.  
XX  
KW TANGO 266; TANGO 261; TANGO 262; TANGO 267;  
KW cellular proliferation; cellular differentiation; cellular adhesion;  
KW von Willebrand factor-associated disorder; cell trafficking; cancer;  
KW hematopoietic associated disease; atelectasis; pulmonary congestion;  
KW edema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;  
KW intestinal disorder; spleen associated disease; renal disorder;  
KW cardiovascular disorder; ischemic heart disease; hydrocephalus;  
KW brain herniation; iatrogenic disease; inflammation; meningitis;  
KW Alzheimer's Disease; cerebral toxoplasmosis; Parkinson's disease;  
KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200052022-A1.  
XX  
XX  
PD 08-SEP-2000.  
XX  
PF 01-MAR-2000; 2000WO-US05226.  
XX  
PR 01-MAR-1999; 99US-0122458.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Barnes TM, Holtzman DA, Sharp JD, Fraser CC;  
XX  
DR WPI; 2000-579269/54.  
DR N-PSDB; AAA75159.  
XX  
PT Novel human and murine secreted proteins designated TANGO 216, 261,  
PT 262, 266 and 267 useful as modulating agents of cellular processes,  
PT e.g. for treating cancer -  
XX  
PS Disclosure; Page -; 175pp; English.  
XX  
CC AAB18455-57 represent human TANGO 216 proteins. The specification also  
CC describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO  
CC polypeptides can be used to modulate cellular proliferation, modulate  
CC cellular differentiation and/or modulate cellular adhesion. The  
CC proteins can be used to treat any von Willebrand factor-associated  
CC disorder, regulate extracellular matrix structuring, cellular adhesion,  
CC and cell trafficking and/or migration, modulate cellular interactions,  
CC modulate cell adhesion in proliferative disorders, such as cancer,  
CC modulate the proliferation, differentiation, and/or function of cells  
CC that appear in the bone marrow, and leukocytes, treat bone marrow, blood  
CC and hematopoietic associated diseases and disorders, atelectasis,  
CC pulmonary congestion or oedema, emphysema, chronic bronchitis, bronchial  
CC asthma and bronchiectasis, intestinal disorders, spleen associated  
CC diseases, modulate renal disorders, treat cardiovascular disorders such

CC as ischemic heart disease, modulate the proliferation, differentiation,  
CC and/or function of bone and cartilage cells and to treat bone and/or  
CC cartilage associated diseases or disorder. They may also be used to  
CC treat disorders associated with the ovaries, and cerebral oedema,  
CC hydrocephalus, brain meningitis, iatrogenic disease, inflammations,  
CC bacterial and viral encephalitis, Alzheimer's disease, cerebral  
CC toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers,  
CC hydrocephalus and encephalitis, and treat hepatic disorders.  
CC note: the present sequence does not appear in the specification; it was  
CC created using information provided.

XX SQ Sequence 488 AA;

Query Match 44.7%; Score 506; DB 21; Length 488;  
Best Local Similarity 53.4%; Pred No. 3.3e-50;  
Matches 101; Conservative 37; Mismatches 49; Indels 2; Gaps 1;  
QY 1 MSFIVFSTRGTTLMKLTEDREQIRQGLELQKVLPGDGYMHGFERASEQIYYENRQCY 60  
Db 80 lsfivfssqatllpltdgrgkiskgledlkrvsvpgetyihbegiklaneqi--qkaggl 137  
QY 61 RTASVLIATDCELHEDLFFYSERANRSRDIGAIYCVGVKDFNETQLARTADSKDHVF 120  
Db 138 kssliialtdgldglvpsyaekaisrlgasvycvvgldfefaqleriadskeqvf 197  
QY 121 PVDGFOALQGIHISILKSCIEILAAEPSTICAGESFOVVYRGNGFRHARNVDRVLCSE 180  
Db 198 pvkggfgalkginsilaqstcelldqpssvcvgeefqivlsgrgfmglngsvlcty 257  
QY 181 KINDSVTLS 189  
Db 258 tvnetytts 266

RESULT 10  
AAB19662  
ID AAB19662 standard; Protein; 587 AA.  
XX  
AC AAB19662;  
XX  
DT 06-DEC-2001 (first entry)  
XX  
DE Human novel extracellular matrix protein, Seq ID NO 312.  
XX  
KW Human; secreted extracellular matrix protein; immunomodulatory;  
KW Anti-HIV; antianemic; antirheumatic; antisclerotic; cardiac; vascular;  
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytostatic;  
KW antialzheimers; immune/autoimmune disease; HIV infection; anaemia;  
KW human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;  
KW cancers; hyperproliferative disorder; breast neoplasm; melanoma;  
KW Sezary syndrome; Gaucher's disease; neurological diseases;  
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;  
KW cardiac arrest; tachycardia; angina; infection; corneal infections;  
KW wound healing; immunogen; gene therapy; antisense; food additive.  
XX  
OS Homo sapiens.  
XX  
PN WO200155368-A1.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01348.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 01-SEP-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2000US-0254097.  
PR 05-JAN-2001; 2000US-0259678.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI: 2001-465572/50.  
N-PSDB; AAS31233.  
Nucleic acid molecules encoding human secreted extracellular matrix proteins, used in preventing, treating or ameliorating a disorder, e.g. Alzheimer's and Parkinson's diseases and cancers -  
Claim 11; SEQ ID No 312; 577pp; English.  
The invention relates to isolated nucleic acid molecules encoding novel human secreted extracellular matrix proteins (SPs). The CC polynucleotides and proteins are used to prevent, treat a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, CC chickens or sheep. For example, disorders associated with decreased

expression of SPs. The SP polynucleotide or a vector expressing them may be administered to treat diseases by gene therapy. Antisense molecules may be administered to down regulate expression of SPs by binding with the cells own genes and preventing their expression. The polynucleotides may also be used as DNA probes in diagnostic assays. The SPs may also be used as antigens to produce antibodies and to identify modulators (agonists and antagonists) of the SPs. The anti-(SP) antibodies and antagonists may also be used to down regulate expression and activity of SP and as diagnostic agents for detecting the presence of SPs in samples. The disorders include for example: immune/autoimmune diseases (e.g. HIV (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis and multiple sclerosis), cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of the breast or liver, Sezary syndrome and Gaucher's disease), neurological diseases (e.g. Alzheimer's disease, Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac arrest, tachycardia and angina), infections caused by bacteria, viruses and fungi and ocular disorders (e.g. corneal infections). Other uses include wound healing, maintenance of organs before transplantation,

Query Match 44.7%; Score 505; DB 22; Length 587;  
Best Local Similarity 53.4%; Pred. No. 5.8e-50;  
Matches 101; Conservative 36; Mismatches 50; Indels 2; Gaps 1;

QY 1 MSFIVSTGTTLMKLTDEQIRQGLELQKVLPGDYMHEGFEFASQIYYENRQGY 60

Db 236 lsfivssqatliilptgdrgkiskgledlkrvspvgetyihgklaneqi--qkaggl 293

QY 61 RTASVITATDGLHEDLFFYSEREANRSRLGAIYVCGVKDFNETQLARTADSKDHVF 120

Db 294 ktsiiaiatdgdkgldglvpsyaekaisrsgasvycvldfeqaqlariadskdqvf 353

QY 121 PVNDGFQALQGIITHSLKSCIEILAAEPSTICAGSFQVVRGNGFRHARNVDRLCSF 180

Db 354 pvkggfgalkgiinsilaqscleilelqpsvscvgeefqivisgrfmglsrngsvlcty 413

QY 181 KINDSVTLSK 189

Db 414 tvnetytts 422

RESULT 11

AAB18448  
ID AAB18448 standard; Protein; 487 AA.

AC AAB18448;

DT 15-JAN-2001 (first entry)

DE Amino acid sequence of a murine TANGO 216 polypeptide.

TANGO 266; TANGO 216;; TANGO 261; TANGO 262; TANGO 267;  
cellular proliferation; cellular differentiation; cellular adhesion;  
von Willebrand factor-associated disorder; cell trafficking; cancer;  
hematopoietic associated disease; atelectasis; pulmonary congestion;  
oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;  
intestinal disorder; spleen associated disease; renal disorder;  
cardiovascular disorder; ischemic heart disease; hydrocephalus;  
brain herniation; iatrogenic disease; inflammation; meningitis;  
Alzheimer's Disease; cerebral toxoplasmosis; Parkinson's disease;  
multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.

Mus sp.

Key Location/Qualifiers

FT Peptide 1..33 /note= "signal peptide"

FT Domain 34..79 /note= "extracellular domain"

FT Protein 34..487 /note= "mature protein"

FT Domain 80..97 /note= "transmembrane domain"

FT Domain 98..317

FT Domain /note= "cytoplasmic domain"  
318..341

FT Domain /note= "transmembrane domain"  
342..487

FT Domain /note= "extracellular domain"

PN WO200052022-A1.

PD 08-SEP-2000.

PF 01-MAR-2000; 2000WO-US05226.

PR 01-MAR-1999; 99US-0122458.

PA (MILL-) MILLENNIUM PHARM INC.

PI Barnes TM, Holtzman DA, Sharp JD, Fraser CC;

DR WPI; 2000-579269/54.

DR N-PSDB; AAA75150.

PT Novel human and murine secreted proteins designated TANGO 216, 261,  
262, 266 and 267 useful as modulating agents of cellular processes,  
e.g. for treating cancer -

PS Claim 8; Fig 2A-C; 175pp; English.

The present sequence represents a murine TANGO 216 polypeptide. The specification also describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO polypeptides can be used to modulate cellular proliferation, modulate cellular differentiation and/or modulate cellular adhesion. The proteins can be used to treat any von Willebrand factor-associated disorder, regulate extracellular matrix structuring, cellular adhesion, and cell trafficking and/or migration, modulate cellular interactions, modulate cell adhesion in proliferative disorders, such as cancer, modulate the proliferation, differentiation, and/or function of cells that appear in the bone marrow, and leukocytes, treat bone marrow, blood and hematopoietic associated diseases and disorders, atelectasis, pulmonary congestion or oedema, emphysema, chronic bronchitis, bronchial asthma and bronchiectasis, intestinal disorders, spleen associated diseases, modulate renal disorders, treat cardiovascular disorders such as ischemic heart disease, modulate the proliferation, differentiation, and/or function of bone and cartilage cells and to treat bone and/or cartilage associated diseases or disorder. They may also be used to treat disorders associated with the ovaries, cerebral oedema, hydrocephalus, brain herniations, iatrogenic disease, inflammations, bacterial and viral meningitis, Alzheimer's disease, cerebral toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers, hydrocephalus and encephalitis, and treat hepatic disorders.

SQ Sequence 487 AA;

Query Match 43.8%; Score 495; DB 21; Length 487;

Best Local Similarity 53.7%; Pred. No. 6.5e-49;

Matches 102; Conservative 32; Mismatches 54; Indels 2; Gaps 1;

QY 1 MSFIVSTGTTLMKLTDEQIRQGLELQKVLPGDYMHEGFEFASQIYYENRQGY 60

Db 80 lsfivssqatliilptgdrgkiskgledlkrvspvgetyihgklaneqi--qkaggl 137

QY 61 RTASVITATDGLHEDLFFYSEREANRSRLGAIYVCGVKDFNETQLARTADSKDHVF 120

Db 138 kassiiiaitdgdkgldglvpsyaekaisrsgasvycvldfeqaqlariadskdqvf 197

QY 121 PVNDGFQALQGIITHSLKSCIEILAAEPSTICAGSFQVVRGNGFRHARNVDRLCSF 180

Db 198 pvkggfgalkgiinsilaqscleilelqpsvscvgeefqivisgrvtsishdgsvlctf 257

QY 181 KINDSVTLSK 190

Db 258 tanstytkse 267

Sequence 487 AA;



KW graft versus-host diseases; rheumatoid arthritis; psoriasis;  
 KW inflammatory bowel disease; septic shock; ulcerative colitis;  
 KW Crohn's disease; chronic myelogenous leukemia; cancer; liver  
 KW disease; Hodgkin's disease; osteoarthritis; Lyme s disease;  
 KW cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;  
 KW systemic lupus erythematosus; transgenic animal; diagnosis;  
 KW prognosis; probiatic; therapeutic; mouse.

OS Mus musculus.

XX

PN WO200039284-A1.

XX

PD '06-JUL-2000.

XX  
75  
73-DEC-1999:

PF 73-DEC-1993,  
YY

PR 30-DEC-1998; 98US-0223546.

XX  
XX

PA (MILL-) MILLENNIUM PHARM INC

XXXX

PI Holtzman DA;

XX  
DB WBT: 2000-465

DR WEI, 2000 403  
DR N-PSDB: AAA47

XX  
XX

PT Novel nucleic

PT 224 and 239 F

PT arthritis, psoriasis and autoimmune diseases  
XX  
PS Claim 8: Fig 27: 209pp: English.

XX

Nucleic acids encoding TANGO polypeptides are useful as modulating agents for regulating cellular processes like asthma, graft versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory bowel disease, septic shock, ulcerative colitis, Crohn's disease, chronic myelogenous leukemia, cancer, liver disease, Hodgkin's disease, osteoarthritis, Lyme's disease, cachexia and autoimmune diseases e.g. myasthenia gravis, autoimmune diabetes and systemic lupus erythematosus. The nucleic acids are also useful for producing transgenic animals and the TANGO polypeptides themselves. Partial TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in forensic biology, for diagnostic assays, prognostic assays, pharmacogenomics and for monitoring clinical trials. TANGO polypeptides are suitable for both prophylactic and therapeutic methods for treating a subject at risk of a disorder or having a disorder associated with aberrant TANGO expression. A wide range of cellular disorders can be treated.

AA	Sequence	381 AA:
S0		

Query Match	36.7%	Score 415;	DB 21;	Length 381;
Best Local Similarity	97.6%	Pred. No. 1.1e-39;		
Matches	81	Mismatches	2	Gaps 0;
Conservative				

Qy 108 QLRIADSKDHVPVNDGFQALQGIHSILKSKCIEILAAEPSTICAGESFQVVVRGNF 167

```

DD      4 qiaiauskunvipvnoygrqiaiqgransnnsbcbraawefstetnfj-1
QY      168 RHARNVDRVLCSEFKINDSVTLK 190
          |||||

```

64 rharnvdrv\cscfkindsvtlne 86

22

Search completed: August 9, 2002, 10:32:15  
Job time: 134 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:33:08 ; Search time 45.48 Seconds  
(without alignments)  
178.842 Million cell updates/sec

Title: US-09-970-076-8  
Perfect score: 1728  
Sequence: 1 MATARRALGIGFQWLSLAT.....TTHCSLHKIASPTTAACNE 333

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	166	9.6	1155	1	US-08-286-889-46
2	166	9.6	1155	1	US-08-485-618-46
3	166	9.6	1155	1	US-08-362-652-46
4	166	9.6	1155	2	US-08-605-672-46
5	166	9.6	1155	2	US-08-482-293A-46
6	166	9.6	1155	2	US-08-943-363-46
7	166	9.6	1155	4	US-09-193-043-46
8	166	9.6	1161	1	US-08-485-618-53
9	166	9.6	1161	2	US-08-362-652-53
10	166	9.6	1161	2	US-08-605-672-53
11	166	9.6	1161	2	US-08-482-293A-53
12	166	9.6	1161	2	US-08-943-363-53
13	166	9.6	1161	4	US-09-193-043-53
14	159.5	9.2	1151	1	US-08-286-889-37
15	159.5	9.2	1151	1	US-08-485-618-37
16	159.5	9.2	1151	1	US-08-362-652-37
17	159.5	9.2	1151	2	US-08-605-672-37
18	159.5	9.2	1151	2	US-08-482-293A-37
19	159.5	9.2	1151	4	US-09-193-043-37
20	159.5	9.2	1151	4	US-09-193-043-37
21	159.5	9.2	1161	1	US-08-173-497-2
22	159.5	9.2	1161	1	US-08-286-889-2
23	159.5	9.2	1161	1	US-08-485-618-2
24	159.5	9.2	1161	1	US-08-485-618-55
25	159.5	9.2	1161	1	US-08-485-618-99
26	159.5	9.2	1161	1	US-08-362-652-2
27	159.5	9.2	1161	1	US-08-362-652-55

28	159.5	9.2	1161	2	US-08-605-672-2	Sequence 2, Appli
29	159.5	9.2	1161	2	US-08-605-672-55	Sequence 55, Appl
30	159.5	9.2	1161	2	US-08-605-672-99	Sequence 99, Appl
31	159.5	9.2	1161	2	US-08-482-293A-2	Sequence 2, Appli
32	159.5	9.2	1161	2	US-08-482-293A-55	Sequence 55, Appl
33	159.5	9.2	1161	2	US-08-482-293A-99	Sequence 99, Appl
34	159.5	9.2	1161	2	US-08-943-363-2	Sequence 2, Appli
35	159.5	9.2	1161	2	US-08-943-363-55	Sequence 55, Appl
36	159.5	9.2	1161	2	US-08-943-363-99	Sequence 99, Appl
37	159.5	9.2	1161	4	US-09-193-043-2	Sequence 2, Appli
38	159.5	9.2	1161	4	US-09-193-043-55	Sequence 55, Appl
39	159.5	9.2	1161	4	US-09-193-043-99	Sequence 99, Appl
40	149.5	8.7	435	5	PCT-US95-04439-1	Sequence 1, Appli
41	145.5	8.4	1152	2	US-08-476-062A-43	Sequence 43, Appl
42	145.5	8.4	1152	5	PCT-US96-01314-43	Sequence 43, Appl
43	145.5	8.4	1152	6	5424399-2	Patent No. 5424399
44	145.5	8.4	1153	1	US-08-173-497-3	Sequence 3, Appli
45	145.5	8.4	1153	1	US-08-286-889-3	Sequence 3, Appli

## ALIGNMENTS

RESULT 1  
US-08-286-889-46  
; Sequence 46, Application US/08286889  
; Patent No. 5470953  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Mich  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/08/286,889  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: P38,659  
; REFERENCE/DOCKET NUMBER: 27866/32168  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-286-889-46

Query Match 9.6%; Score 166; DB 1; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 1.9e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;  
QY 37 PACYG-GFDLYFILDKSGSV-LHHWNIYFVEQLAHKFIQPLRMSFIVSTRGTLTKM 94

Db	144	PECPQGMDFIAFLIDGSGSIDQSDFTOMKDFKALMGQLASTSTFSLMQYENILKTHEFT	203
QY	95	LTEDR-----EQTRQGLEELQKVLPGDDYMHGFFERASEQIYVENQRGYRTA-SVIAAL	148
Db	204	FTEFKSSLSPQSVDIAIVQIQ-----GLTVTAGIOKVVKLEFHSKNGARKSAKKILIVI	258
QY	149	TDGELHEDLFYSE--REANRSRDLAGIVYCVGVKD-FNE-TOLARI-----ADSKDHVF	199
Db	259	TDGKFRDPLRYHRVPEAKA---GIIRYAIGVDFAFREPTALQELNTITGSAPSDHVF	315
QY	200	PVNDGFOALQGIITHSLKSCIELAEPSTICAGESFQVVVRGNGFRHARNVD	253
Db	316	KVGN-FVALRSIQRIQEK-----IFATEGTESRSSSSQHEMSQEGFSSALSD	364

```

RESULT      2
US-08-485-618-46
; Sequence 46, Application US/08485618
; Patent NO. 5728533
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5728533e1 Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,618
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32797
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1155 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-485-618-46

```

Query Match 9.6%; Score 166; DB 1; Length 1155;  
 Best Local Similarity 28.2%; Pred. No. 1.9e-09;  
 Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;  
 QY 37 PACYG-GFDLYFILDKSSV-IHHWNEIYYFVEQLAHKFI SPQLRFSFVFSTRGTTLMK 94  
 | | | | | : : : | | | : : | | | : : |

```

Db      144  PCCPGQMDIAPLDGSGSIDQSDFTQMKDFVVKALMGQLASTSTSFSLMQYSNLIKTHFT 203
QY      95  LTRDR-----EOIRGLEELQKVLPGGDTYMHGCFERASEQIYYENRGYRTA-SVIAL 148
Db      204  FTEFKSLSPQSLVDAIVQLQ-----GLTVTASGIQKVVKELFHSKNGARKSAKKILIVI 258
QY      149  TDGELHEDLFFYSE--REANKSRDLGAIVCVGVKQD-FNE-TQLARI-----ADSKDHVF 199
Db      259  TDGOKFRDPLEYRHVPAEKA---GIIRVAIGVGDFAFRPTALQELNTIGSAPSQDHVF 315
QY      200  PVNDGFOALQGIHSLKKSCTEILAAEPSTICAGESFQVVVGRNGFRHARNVD 253
Db      316  KVGN-FVALRSQRQIQEK-----IFAIEGTESRSSSFQHEMSQEGFSALSMD 364

RESULT 3
US-08-362-652-46
; Sequence 46, Application US/08362652
; Patent No. 5766850
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,652
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32391
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1155 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-362-652-46

```

	Query Match	9.68;	Score 166;	DB 1:	Length 1155;	
	Best Local Similarity	28.2%;	Pred. No. 1.9e-09;			
	Matches	66;	Conservative	42;	Mismatches 96;	Indels 30; Gaps 12;
QY	37 PACYG-GPDLYFILDKSGSV-LHHWNEIYFVEQLAHKFTSPOLRMFSVFSTRGTTLMK	94	:   :   :   :   :   :			
Dd	144 PECGGEMDIAFLDGSGSIDOSFOFMKDFFVKALMGQLASTSFSLMOYNILKTHT	203	:   :   :   :   :   :			
QY	95 LTEDR-----EQIRGLDEELOKVLPFGDTTYMHEGFASERAGSIYYENRGQVRGA-SVIIAL	148	: : : :   :     : :   : : : :			

Db 204 FTEFKSSLSQSLVDAIVQLQ-----GLUTYASGIQKVVVKELFHSKNGARKSAKKILIVI 258  
Qy 149 TDGELHEDLFYSE--REANRSRDGLAIVYCVGVKD-FNE-TQLARI-----ADSKDHVF 199  
Db 259 TDGQKFRDPLRYRHVIPAEGA---GIIRYAIGVDAPREPTALOELNTIGSAPSQDHVF 315  
Qy 200 PVNDQFALQGIHIIILKKSCEILAAEPSTICAGESQVYVVRNGFRHARNVD 253  
Db 316 KVG-N-FVALRSIQRIQEK-----IFAIBGTESRSSSPQHMSQEGFSSALSMD 364

RESULT 4  
US-08-605-672-46  
; Sequence 46, Application US/08605672  
; Patent No. 5817515  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: NO. 5817515el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/605,672  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32684  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-605-672-46

Query Match 9.6%; Score 166; DB 2; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 1.9e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;  
Qy 37 PACYG-GFDLVFILDKSGSV-LHHWNEIYFYVEQLAHFISPOLRMSFVSTRTGTTLMK 94  
Db 144 PCPCQEMDIAFLIDSGSIDQSDFTQMKDFVKALMGQLASTSTFSLSMQYSNLIKTHFT 203  
Qy 95 LTEDR-----EQIROGLEELQVLPGGDTYMHGFEFRASEQIYYENROGYRTA-SVIAL 148  
Db 204 FTEFKSSLSQSLVDAIVQLQ-----GLUTYASGIQKVVVKELFHSKNGARKSAKKILIVI 258

Qy 149 TDGELHEDLFYSE--REANRSRDGLAIVYCVGVKD-FNE-TQLARI-----ADSKDHVF 199  
Db 259 TDGQKFRDPLRYRHVIPAEGA---GIIRYAIGVDAPREPTALOELNTIGSAPSQDHVF 315  
Qy 200 PVNDQFALQGIHIIILKKSCEILAAEPSTICAGESQVYVVRNGFRHARNVD 253  
Db 316 KVG-N-FVALRSIQRIQEK-----IFAIBGTESRSSSPQHMSQEGFSSALSMD 364  
RESULT 5  
US-08-482-293A-46  
; Sequence 46, Application US/08482293A  
; Patent No. 5831029  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: NO. 5831029el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/482,293A  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32684  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-482-293A-46

Query Match 9.6%; Score 166; DB 2; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 1.9e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;  
Qy 37 PACYG-GFDLVFILDKSGSV-LHHWNEIYFYVEQLAHFISPOLRMSFVSTRTGTTLMK 94  
Db 144 PCPCQEMDIAFLIDSGSIDQSDFTQMKDFVKALMGQLASTSTFSLSMQYSNLIKTHFT 203  
Qy 95 LTEDR-----EQIROGLEELQVLPGGDTYMHGFEFRASEQIYYENROGYRTA-SVIAL 148  
Db 204 FTEFKSSLSQSLVDAIVQLQ-----GLUTYASGIQKVVVKELFHSKNGARKSAKKILIVI 258

us-09-970-076-8.ra1

Fri Aug 9 10:57:07 2002

QY 149 TDGELHEDLFFYSE--REANRSDLGAIYVGVKDFNE-TOLARI-----ADSKDHVF 199  
 Db 259 TDGQKFRDPLEYRHVPIEAEKA--GIIRYAGVGDFAREPTALQELNTIGSAPSQDHVF 315  
 QY 200 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQVVRGNGFRHARNVD 253  
 Db 316 KVGNFVALRSIQRIQIEK-----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

RESULT 6  
 US-08-943-363-46  
 ; Sequence 46, Application US/08943363  
 ; Patent No. 5837478  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gallatin, W. Michael  
 ; APPLICANT: Van der Vieren, Monica  
 ; TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit  
 ; NUMBER OF SEQUENCES: 114  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 233 South Wacker Drive, 6300 Sear Tower  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA: US/08/943,363  
 ; FILING DATE:  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/173,497  
 ; FILING DATE: 23-DEC-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/286,889  
 ; FILING DATE: 5-AUG-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/362,652  
 ; FILING DATE: 21-DEC-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Williams Jr., Joseph A.  
 ; REGISTRATION NUMBER: 38,659  
 ; REFERENCE/DOCKET NUMBER: 27866/32684  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312-474-6300  
 ; TELEFAX: 312-474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 46:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1155 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-943-363-46

Query Match 9.6%; Score 166; DB 2; Length 1155;  
 Best Local Similarity 28.2%; Pred. No. 1.9e-09;  
 Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;  
 QY 37 PACYV-GFDLYFILDKGSV-LHHWNEIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMK 94  
 Db 144 PECQEMDI AFLIDGSGSDQSDFTQMKDFVKALMGQLASTSTFSLMOYSNLTHT 203  
 QY 95 LTEDR-----EQIRQGLEELQKVLPGDVTYMHGFEFRAEQIYYENRQGYRTA-SVIAL 148  
 Db 204 FTEPKSSLPQSLVDAIVQLQ-----GLTYASGIQVKVKELFHSGKNGARKSAKILIVI 258  
 QY 149 TDGELHEDLFFYSE--REANRSDLGAIYVGVKDFNE-TOLARI-----ADSKDHVF 199

Db 259 TDGQKFRDPLEYRHVPIEAEKA--GIIRYAGVGDFAREPTALQELNTIGSAPSQDHVF 315  
 QY 200 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQVVRGNGFRHARNVD 253  
 Db 316 KVGNFVALRSIQRIQIEK-----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

RESULT 7  
 US-09-193-043-46  
 ; Sequence 46, Application US/09193043  
 ; Patent No. 6251395  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gallatin, W. Michael  
 ; APPLICANT: Van der Vieren, Monica  
 ; TITLE OF INVENTION: No. 6251395el Human 2  
 ; FILE REFERENCE: 27866/35004  
 ; CURRENT APPLICATION NUMBER: US/09/193,043  
 ; CURRENT FILING DATE: 1998-11-16  
 ; EARLIER APPLICATION NUMBER: 08/173,497  
 ; EARLIER FILING DATE: 1993-12-23  
 ; EARLIER APPLICATION NUMBER: 08/286,889  
 ; EARLIER FILING DATE: 1994-08-05  
 ; EARLIER APPLICATION NUMBER: 08/362,652  
 ; EARLIER FILING DATE: 1994-12-21  
 ; EARLIER APPLICATION NUMBER: 08/943,363  
 ; EARLIER FILING DATE: 1997-10-03  
 ; NUMBER OF SEQ ID NOS: 114  
 ; SOFTWARE: Patent in Ver. 2.0  
 ; SEQ ID NO 46  
 ; LENGTH: 1155  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 ; US-09-193-043-46

Query Match 9.6%; Score 166; DB 4; Length 1155;  
 Best Local Similarity 28.2%; Pred. No. 1.9e-09;  
 Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;  
 QY 37 PACYV-GFDLYFILDKGSV-LHHWNEIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMK 94  
 Db 144 PECQEMDI AFLIDGSGSDQSDFTQMKDFVKALMGQLASTSTFSLMOYSNLTHT 203  
 QY 95 LTEDR-----EQIRQGLEELQKVLPGDVTYMHGFEFRAEQIYYENRQGYRTA-SVIAL 148  
 Db 204 FTEPKSSLPQSLVDAIVQLQ-----GLTYASGIQVKVKELFHSGKNGARKSAKILIVI 258  
 QY 149 TDGELHEDLFFYSE--REANRSDLGAIYVGVKDFNE-TOLARI-----ADSKDHVF 199  
 Db 259 TDGQKFRDPLEYRHVPIEAEKA--GIIRYAGVGDFAREPTALQELNTIGSAPSQDHVF 315  
 QY 200 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQVVRGNGFRHARNVD 253  
 Db 316 KVGNFVALRSIQRIQIEK-----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

RESULT 8  
 US-08-485-618-53  
 ; Sequence 53, Application US/08485618  
 ; Patent No. 5728533  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gallatin, W. Michael  
 ; APPLICANT: Van der Vieren, Monica  
 ; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit  
 ; NUMBER OF SEQUENCES: 103  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 233 South Wacker Drive, 6300 Sear Tower  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States  
 ; ZIP: 60606-6402



Query Match	9.6%	Score 166;	DB 2;	Length 1161;
Best Local Similarity	28.2%	Pred. No. 1.9e-09;		
		Indels 3		

RESULT 12

US-08-943-363-53  
; Sequence 53, Application US/08943363  
; Patent No. 5837478

```

1 GENERAL INFORMATION:
2
3 APPLICANT: Gallatin, W. Michael
4 APPLICANT: Van der Vieren, Monica
5 TITLE OF INVENTION: NO. 5837478e1 Human 2 Integrin Alpha Subunit
6
7 NUMBER OF SEQUENCES: 114
8
9 CORRESPONDENCE ADDRESS:
10 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
11 STREET: 233 South Wacker Drive, 6300 Sear Tower
12 CITY: Chicago
13 STATE: Illinois
14 COUNTRY: United States
15
16 ZIP: 60606-6402
17
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Floppy disk
20 COMPUTER: IBM PC compatible
21 OPERATING SYSTEM: PC-DOS/MS-DOS
22 SOFTWARE: Patent In Release #1.0, Version #1.25
23
24 CURRENT APPLICATION DATA:
25 APPLICATION NUMBER: US/08/943,363

```

Query Match 9.6%; Score 166; DB 2; Length 1161;  
Best Local Similarity 28.2%; Pred. No. 1.9e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 3

[illegible]

RESULT 11  
US-08-482-293A-53  
; Sequence 53, Application US/08482293A  
; Patent No. 5831029

```

GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

```

```
;
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-943-363-53

Query Match          9.6%; Score 166; DB 2; Length 1161;
Best Local Similarity 28.2%; Pred. No. 1.9e-09;
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 37 PACYV-GFDLYFILDKSGSV-LHHWNEIYFVEOLAHKFISPOLRMSFIVFSTRGTILMK 94
DB 144 PECGQEMDIAFLIDSGSIDQSDFTKDKFVKALMGQLASTSTSFSLMQYSNLIKTHFT 203
QY 95 LTEDR-----EQIRQGLEELQKVLPGDGYMHGFEFRASEQIYYENRQGYRTA-SVIAL 148
DB 204 FTEFKSSLSQSLVDIAIVQLQ-----GLVTASGIQKVKELPHSKNGARKSAKKILIVI 258
QY 149 TDGELHEDLFFYSE--REANRSRLGAIVYCVGVKD-FNE-TQLARI-----ADSKDHVF 199
DB 259 TDGQKFRDPLEYRHVPIPEAKA---GIIRYAIGVDGDAFREPTALQELNTIGSAPSODHVF 315
QY 200 PVNDGFOALOGIIHSILKSKCIEILAAEPSTICAGESFOVVVRGNGFRHARNVD 253
DB 316 KVG-N-FVALRSIQRIQIEK----IPAEICTESRSSSFQHEMSQEGFSSALSMD 364

RESULT 13
US-09-193-043-53
; Sequence 53, Application US/09193043
; Patent No. 6251395
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 6251395el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/193,043
; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173,497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286,889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362,652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943,363
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
; LENGTH: 1161
; TYPE: PRT
```

```
;
; ORGANISM: Mus musculus
; US-09-193-043-53

Query Match          9.6%; Score 166; DB 4; Length 1161;
Best Local Similarity 28.2%; Pred. No. 1.9e-09;
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 37 PACYV-GFDLYFILDKSGSV-LHHWNEIYFVEOLAHKFISPOLRMSFIVFSTRGTILMK 94
DB 144 PECGQEMDIAFLIDSGSIDQSDFTKDKFVKALMGQLASTSTSFSLMQYSNLIKTHFT 203
QY 95 LTEDR-----EQIRQGLEELQKVLPGDGYMHGFEFRASEQIYYENRQGYRTA-SVIAL 148
DB 204 FTEFKSSLSQSLVDIAIVQLQ-----GLVTASGIQKVKELPHSKNGARKSAKKILIVI 258
QY 149 TDGELHEDLFFYSE--REANRSRLGAIVYCVGVKD-FNE-TQLARI-----ADSKDHVF 199
DB 259 TDGQKFRDPLEYRHVPIPEAKA---GIIRYAIGVDGDAFREPTALQELNTIGSAPSODHVF 315
QY 200 PVNDGFOALOGIIHSILKSKCIEILAAEPSTICAGESFOVVVRGNGFRHARNVD 253
DB 316 KVG-N-FVALRSIQRIQIEK----IPAEICTESRSSSFQHEMSQEGFSSALSMD 364

RESULT 14
US-08-286-889-37
; Sequence 37, Application US/08286889
; Patent No. 5470953
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Mich
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,889
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32168
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-286-889-37

Query Match          9.2%; Score 159.5; DB 1; Length 1151;
Best Local Similarity 27.9%; Pred. No. 1e-08;
Matches 63; Conservative 41; Mismatches 93; Indels 29; Gaps 11;
```







**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: August 9, 2002, 10:46:25 ; Search time 102.68 seconds  
(without alignments)  
561.037 Million cell updates/sec

Title: US-09-970-076-8

Perfect score: 1728

Sequence: 1 MATAERRALGIGFQWLSLAT.....TTHCSLHKIASGPTTAACME 333

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1649	95.4	368	4 Q96P02	Q96P02 homo sapien
2	582	33.7	245	4 Q96NC7	Q96NC7 homo sapien
3	467	27.0	97	4 Q96EC6	Q96EC6 homo sapien
4	159.5	9.2	1161	11 Q9QVE7	Q9QVE7 rattus norv
5	153.5	8.9	1332	5 Q9BP08	Q9BP08 halocynthia
6	150.5	8.7	724	5 Q04588	Q04588 eimeria max
7	137.5	8.0	712	5 Q43981	Q43981 eimeria ten
8	134	7.8	760	11 Q70350	Q70350 mus musculus
9	132.5	7.7	765	5 Q9U8J9	Q9U8J9 neospora ca
10	131.5	7.6	517	4 Q43853	Q43853 homo sapien
11	131	7.6	660	4 Q9UGC3	Q9UGC3 homo sapien
12	130.5	7.6	920	6 Q28984	Q28984 sus scrofa
13	128	7.4	3767	5 Q9UAI3	Q9UAI3 caenorhabdi
14	123	7.1	1151	11 Q9J130	Q9J130 rattus norv
15	121	7.0	221	11 Q99L17	Q99L17 mus musculus
16	118.5	6.9	3567	11 Q9ES77	Q9ES77 mus musculus

17	117.5	6.8	1140	4 Q9P218	Q9P218 homo sapien
18	117.5	6.8	1207	4 Q9BQ07	Q9BQ07 homo sapien
19	117.5	6.8	2944	11 Q63870	Q63870 mus musculus
20	116	6.7	2104	5 Q21281	Q21281 caenorhabdi
21	116	6.7	2104	5 Q964N4	Q964N4 caenorhabdi
22	114	6.6	937	4 Q96FT5	Q96FT5 homo sapien
23	113.5	6.6	755	4 Q00261	Q00261 homo sapien
24	113.5	6.6	764	4 Q96HX6	Q96HX6 homo sapien
25	112	6.5	650	11 Q9CY21	Q9CY21 mus musculus
26	111	6.4	1905	5 Q9XTPE	Q9XTPE plasmodium
27	110.5	6.4	833	13 Q9YIC5	Q9YIC5 cyprinus ca
28	109.5	6.3	704	10 Q9FF49	Q9FF49 arabidopsis
29	109	6.3	1253	6 Q97566	Q97566 canis fami
30	108.5	6.3	956	11 Q99K64	Q99K64 mus musculus
31	106.5	6.2	562	5 Q01510	Q01510 plasmodium
32	106.5	6.2	747	13 Q91900	Q91900 xenopus lae
33	106	6.1	415	11 Q923K3	Q923K3 mus musculus
34	105.5	6.1	567	5 Q18048	Q18048 caenorhabdi
35	104.5	6.0	769	5 Q00816	Q00816 toxoplasma
36	104.5	6.0	1169	11 Q9QXH4	Q9QXH4 mus musculus
37	104	6.0	205	11 Q63001	Q63001 rattus norv
38	102.5	5.9	537	4 Q96AA0	Q96AA0 homo sapien
39	102	5.9	614	5 Q94674	Q94674 plasmodium
40	102	5.9	714	10 Q948G7	Q948G7 oryza sativ
41	102	5.9	759	4 Q96C67	Q96C67 homo sapien
42	102	5.9	833	5 Q96442	Q96442 strongyloce
43	101.5	5.9	425	5 Q9GZF5	Q9GZF5 caenorhabdi
44	101.5	5.9	559	5 Q01506	Q01506 plasmodium
45	101.5	5.9	559	5 Q01508	Q01508 plasmodium

## ALIGNMENTS

RESULT 1

Q96P02 ID Q96P02 PRELIMINARY; PRT; 368 AA.

AC Q96P02;

DT 01-DEC-2001 (TREMBlrel. 19, Created)

DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)

DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE ANTHRAX TOXIN RECEPTOR.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_Taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Bradley K.A., Moggridge J., Mourez M., Collier R.J., Young J.A.T.;

RT "Identification of the Cellular Receptor for Anthrax Toxin.";

RL Nature 414:0-0(2001).

DR EMBL; AF421380; AAL26496.1; -.

KW Receptor.

SQ SEQUENCE 368 AA; 41157 MW; 8A87B13FFA7D8753 CRC64;

Query Match 95.4%; Score 1649; DB 4; Length 368;  
Best Local Similarity 100.0%; Pred. No. 2e-134;  
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATAERRALGIGFQWLSLATLVLCAGQGRRDGGPACYGFDLYFILDKSGSVLHHN 60

Db 1 MATAERRALGIGFQWLSLATLVLCAGQGRRDGGPACYGFDLYFILDKSGSVLHHN 60

Qy 61 EITYFVEQLAHKFIQPLRMSFIVSTRTTLMKLTEDREQIRQGLEELQKVLPGDDTYM 120

Db 61 EITYFVEQLAHKFIQPLRMSFIVSTRTTLMKLTEDREQIRQGLEELQKVLPGDDTYM 120

Qy 121 HEGFERASEQIYYENRQGYRTASVIALDGHEDLFFYSREANRSLDGAIVYCVG 180

Db 121 HEGFERASEQIYYENRQGYRTASVIALDGHEDLFFYSREANRSLDGAIVYCVG 180

Qy 181 KDFNETQLARIADSKDHVFPVNDGFOALQGIITHSLKSCIEILAAEPSTICAGESFQV 240



```
Db 212 DPQSLVDPIVLQ-----GLTYTATGIRTVMEELFHSKNGSKSAKILLVITDQKQYRD 266
QY 157 LFFYSE--REANRDLGAIYVCVKD-FNE-TOLARI-----ADSKDHVPVNDGFOA 207
Db 267 PLEYSDVTPADKA---GIIRYAIGVGDAFQEPALKELNTIGSAPPQDHVFKVGN-FAA 322
QY 208 LQGIISILKSKICIBIAAEPSTICAGSFQVVRGNGFRHARNVD 253
Db 323 LRSIQRLQLEK----IFAELGTQSRSSSSSFQHEMSQEGFSSALTS 364

RESULT 5
Q9BPQ8 PRELIMINARY; PRT; 1332 AA.
AC Q9BPQ8;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE INTEGRIN ALPHA HRI PRECURSOR.
GN HRTGAL.
OS Halocynthia roretzi (Sea squirt).
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
OC Stolidobranchia; Pyuridae; Halocynthia.
ON NCBI_TaxID=7729;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEMOCYTE;
RX MEDLINE=21103187; PubMed=11160215;
RA Miyazawa S., Azumi K., Nonaka M.;
RT "Cloning and characterization of integrin alpha subunits from the
RT solitary ascidian, Halocynthia roretzi."
RL J. Immunol. 166:1710-1715(2001).
DR EMBL; AB048261; BAB21479.1; -.
DR HSP; P11215; I48X.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VMFA.
DR Pfam; PF01839; FG-GAP; 5.
DR Pfam; PF00357; Integrin_A; 1.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VMFA; 1.
KW Signal.
FT SIGNAL.
FT CHAIN.
SQ SEQUENCE 1332 AA; 145852 MW; 0D9108D2B05CFFAE CRC64;

Query Match 8.9%; Score 153.5; DB 5; Length 1332;
Best Local Similarity 24.1%; Pred. No. 0.00025;
Matches 63; Conservative 42; Mismatches 81; Indels 75; Gaps 10;

QY 29 GGRREDGG-----PACYGGFDLYFILDKSGYLHHWNEIYFEVQLAHKFISPLRMSP 83.
Db 188 GNRKEGSNTECPSS--SGVDVFLVDGSGVGKFNKDKVKNIT----- 232
QY 84 VFSTGTTLMKLTEDREQLELQKVLPG-----GDTYMHGFERAS 128
Db 233 -----AKLDIGKEIVRVGVQVSHYVEGKSKINKQKIITTEISIGEFKLLDNFENAV 283
QY 129 EOIYYENROGYRT-----ASVIALTDGELHEDLFFYSEREANRS 168
Db 284 DRI---QLQGYTTYGRALQKIVRFDADDAYIGNKVQLLLTLDGQAKDNKLIIP--NANKL 338
QY 169 RDLGAIYVCVGVKDFNETOLARIA---DSKDHVFPVNDGFOALQGIISILKSKICIEILA 225
Db 339 RNKGATPAVGVEYDISELKLIASCTDSTDTRVFTVD-FGELDSIVKSLQTEIQSFVLE 397
QY 226 AEPSTICAGSFQVVRNGNF 246
Db 398 GGSARTAG--YEMHFGENG 416
```

```
RESULT 6
Q04588 PRELIMINARY; PRT; 724 AA.
ID Q04588;
AC Q04588;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MAJOR ANTIGEN HOMOLOGOUS SEQUENCE (EMP100).
OS Eimeria maxima.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
OC Eimeria.
ON NCBI_TaxID=5804;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93149203; PubMed=8426611;
RA Pasamontes L.E., Hug D., Huembellin M., Weber G.;
RT "Sequence of a major Eimeria maxima antigen homologous to the Eimeria
RT tenella microneme protein Etp100."
RL Mol. Biochem. Parasitol. 57:171-174(1993).
DR EMBL; M99058; AAA29076.1; -.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR002035; VMFA.
DR Pfam; PF00090; tsp_1; 6.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00209; TSPI; 6.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50092; TSPI; 5.
DR PROSITE; PS50234; VMFA; 1.
SQ SEQUENCE 724 AA; 75808 MW; AC2A0E7A346A7E9E CRC64;
```

```
Query Match 8.7%; Score 150.5; DB 5; Length 724;
Best Local Similarity 25.9%; Pred. No. 0.0002;
Matches 69; Conservative 45; Mismatches 109; Indels 43; Gaps 17;

QY 1 MATAERRALGIGFQWLSTLATLVLCAG-----QGRRREDGGPACVGGFDLYFILDKSGS 54
Db 1 MALLPTQLAPG-W-ALSLVFLAAGLTFHSHAAASSEADQVCTRLLDVLMVVDSEGS 57
QY 55 V-LHHWNEIYFEVQLAHKF-ISP-QLRMSFIVFSTRTTLMKLTEDREQIROGLELQK 111
Db 58 ICTSNVGVKVSFISNFACTMPLSPDDVRVGLVTFGTSAVTRWDLSDSRAQNADLLAAAK 117
QY 112 VLP--GGDTYMHGFERASEQIYYENRQYR--TASVIALTDGELHEDLFFYSERE--- 164
Db 118 KLPYAAGSTYTHLGLAKA-EELIFSQKGRDNAPKMLVMTDGA-----SSRRSQ 168
QY 165 ---ANRSRDLGAIYVCVGV-KDFNETOLARIA--DSKDHV-FP--VNDGFOALQGIHSI 215
Db 169 LSAAEKLNRGVIIVVVGVTGVNSAECSIAACDTSIDTVECPRYLQSNWGVSSQINGI 228
QY 216 LKSCIEI----LAAPSTI--CAGE 235
Db 229 IRAACKDLAKDAVCSEWSEYGPCEGE 254

RESULT 7
Q03981 PRELIMINARY; PRT; 712 AA.
ID Q03981;
AC Q03981;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MICRONEME PROTEIN ETMIC-1 PRECURSOR.
GN MIC-1.
OS Eimeria tenella.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
OC Eimeria.
ON NCBI_TaxID=5802;
RN [1]
```

Fri Aug 9 10:57:09 2002

RA Rowen L., Madan A., Qin S., Hall J., Dahl T., James R., Dickhoff R.,  
 RA Schaffer T., Ratcliffe A., Abbasi N., Loretz C., Lasky S., Hood L.;  
 RT "Sequence of the mouse MHC class III region."  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 291-760 FROM N.A.  
 RA Rowen L., Qin S., Lasky S.R., Loretz C., Dors M., Mahairas G.,  
 RA Hood L.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BREAST TUMOR;  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC TRYPsin FAMILY.

RA EMBL; AF109906; RAC84162.1; -;  
 DR EMBL; AF049850; AAC05284.1; -;  
 DR EMBL; BC011086; AAH11086.1; -;  
 DR HSSP; P00761; 1EPT.  
 DR MGD; MGI:88226; C2.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR000436; Sushi\_SCR\_CCP.  
 DR InterPro: IPR001254; Trypsin.  
 DR InterPro: IPR002035; WFA.  
 DR Pfam; PF00084; sushi; 2.  
 DR Pfam; PF00092; vwa; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PRINTS; PR00453; WFADOMAIN.  
 DR SMART; SM00032; CCP; 2.  
 DR SMART; SM00020; Tryp\_SPC; 1.  
 DR SMART; SM00327; WVA; 1.  
 DR PROSITE; PS0240; TRYPsin\_DOM; 2.  
 DR PROSITE; PS00134; TRYPsin\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00135; TRYPsin\_SER; UNKNOWN\_1.  
 DR PROSITE; PS0234; WFA; 2.  
 DR Hydrolase; Serine protease.  
 KW SIGNAL.  
 FT CHAIN 37 712  
 FT SEQUENCE 712 AA; 74777 MW; 15B8F3C190B70F73 CRC64;

Query Match 8.0%; Score 137.5; DB 5; Length 712;  
 Best Local Similarity 24.7%; Pred. No. 0.0025;  
 Matches 67; Conservative 43; Mismatches 108; Indels 53; Gaps 14;

QY 1 MATAERALGIGFQWLATLVLCAG-----OGGRREDGG--PACYGGFDLYFLDK 51  
 DB 1 MAPLPRLRLA---PCRLSLLVGLLAASFAFSLQPGATTSSGQDVCTSLDVMVVD 57  
 QY 52 SGSV-LHHNWEIYFVEQLAHK-ISPQ-LRMSFVIFSTRGTLMLKLTEDREQIRGLEE 108  
 DB 58 SGSGTSTNPKRQFIEDFVNSMPISPDVVRVGLITFAT-----RSKVRNLS 106  
 QY 109 LQKVLPG-----GDTYMEHGFERASEQIYENROGY--TASVIAITDGL 153  
 DB 107 PKATNPISATSAARSUSYSTGVTHYGLQDA-KLLYDTNAGARNVPLVMTDGA- 164  
 QY 154 HEDLFFYSEREANRDLGAIVYCVGY-KDFNETQIARIADSKDHVFP-----VNDGFQ 206  
 DB 165 -SNLPQTRSSAARLDAGAIIVVLGVGVNSSECRSIAGCSTSNCPRYLQSNWSNV 223  
 QY 207 ALQGIHIIILKSCIEILAEPTI--CAGE 235  
 DB 224 QVNGIIRAKACKDLAKDAVCSEWSEYGPVGE 254

RESULT 8  
 ID O70350 PRELIMINARY; PRT; 760 AA.  
 AC O70350;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1999 (TrEMBLrel. 11, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE COMPLEMENT FACTOR C2 (COMPLEMENT COMPONENT 2) (WITHIN H-2S).  
 GN C2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-129;

Query Match 7.8%; Score 134; DB 11; Length 760;  
 Best Local Similarity 23.5%; Pred. No. 0.0056;  
 Matches 75; Conservative 55; Mismatches 115; Indels 74; Gaps 16;

QY 41 GGFDLIFLDKSGVLLHWNIEY-YFVEQLAHKISPOLRMS--FIVFSTRGTMLKLT 97  
 DB 258 GHLNLYLLDASQSVTEKDFDFKKSALMVERIFSEVNVSVAIITFASQPKTMSILS 317  
 QY 98 DREQIRQGLELEQKVLPGDVTYMHGFERASEQIYEE-----NRQGYRTAS- 143  
 DB 318 ERSQ-----DVTEVITSLDSASYKDHENATGTNTYEVLRVYSMMOSQMDRLGME 371  
 QY 144 -----VIATDGEHEDLFFYSEREANRDLGAI-----VYCVGY-----KD 182  
 DB 372 KEIRHTIILLTDGK--SNMGSPKKAIVTRIRKLLSIEGRDDYLDIYAIGVKLDVWKE 429  
 QY 183 FNETOLARIADSKDHVFPVNDGFQALQGIHIIILKSCIEILAEPTICAGESFQVVR 242  
 DB 430 LNE--LGSKKDGERRHAFILODA-KALQQLFEHMLDVSKL-----TDTICG-----V 472  
 QY 243 GNGFRHARNVDRV--LCSPKINDSVTLNEKPFVSEDTYLLCPAPILKEVGK--AALQVS 298  
 DB 473 GNMASANASQERTPWQVTFKPKSKETCOGS--LISQDWLVLTAAHCFHDIQMEDHHLWRV 530  
 QY 299 MND----GLSFISSSVII 312  
 DB 531 VGDPTSQHGKEFLVEDVII 549

RESULT 9  
 QY08J9  
 ID QY08J9 PRELIMINARY; PRT; 765 AA.



Fri Aug 9 10:57:09 2002

us-09-970-076-8.isrpt

QY 44 DLYFILDKSGSV-LHHWNEIYVFEQLAHKF-ISP-QLRMSFIVEST---RGTTIMLKUTE 97  
Db 469 DIVFLVDGSGYSGIANFYKVRFLVLYKVSFEISPNRQVLSILVOYSRDPHTFEFLKFKTK 528  
QY 98 DREQIROGLEELQKVLVPG--GGDTYMHGEGFERASEQIYYENRQGYET--ASVIALTLTGEL 153  
Db 529 VEDII-----EAINTPYRGSGTNGKAMTYVREKIFVPSK-GSRNVPKVMILITDKG- 581  
QY 154 HEDLFYISERENRSDLGAIVYCVGKDFNETQLARIAD--SKDHVPVNDGFOALQGI 211  
Db 582 SSDAF---RDPALIKLRSDVEIFANGVCKDAVRSELEAIASPPAETHVFTVED-FDAFQRI 637  
QY 212 IHSILKKSCTIE---LAA 226  
Db 638 SFELTQSICLRIEQELAA 655  
RESULT 12  
Q28984 ID Q28984 PRELIMINARY; PRT; 920 AA.  
AC Q28984;  
DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
DT 01-FEB-1997 (T-EMBLrel. 02, Last sequence update)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
DE CD11B (FRAGMENT).  
DN GN  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Lee J.-K., Schock L.B., Rutherford M.S.;  
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U40072; AAB16869.1; -.  
DR HSSP: P11215; IAHX.  
DR InterPro: IPR000413; Integrin\_alpha.  
DR InterPro: IPR002035; VWFA.  
DR Pfam: PF01839; FG-GAP; 4.  
DR Pfam: PF00092; vwa; 1.  
DR PRINTS: PR01185; INTEGRINA.  
DR PRINTS: PR00453; VWFADOMAIN.  
DR SMART: SM00191; Int\_alpha; 4.  
DR SMART: SM00327; VWA; 1.  
DR PROSITE: PS50234; VWFA; 1.  
FT NON\_TER 1  
FT NON\_TER 920 920  
FT SEQUENCE 920 AA; 102440 MW; E96CC51E350DD5AC CRC64;

Query Match 7.6%; Score 130.5; DB 6; Length 920;  
Best Local Similarity 26.4%; Pred. No. 0.015;  
Matches 63; Conservative 41; Mismatches 72; Indels 63; Gaps 15;

QY 44 DLYFILDKSGSVLHHWNEIYVFEQLAHKFISP---QLRMSFIVESTRGTTIMLKUTED-- 98  
Db 17 DIAFLIDGSGSI---NRLDF---QRMKEFVTVMGQFQSKTLFA-----LMQYSEDFY 64  
QY 99 -----REQIRQGLEELQKVLVPGDGYTMHGEFERASEQIYYENRQGYR- 140  
Db 65 THFTFNDFRNRPSPKLLVPRIKQLL-----GRTHATGIRKVVRELPH-SKSGARE 114  
QY 141 -TASVIALTDGELHEDLFFYSE--REANRSDLGAIVYCVGVKD-FN-----ETOLARIA 192  
Db 115 NALKILVITDGEKFGDPLGVEDVPEADRK---GVIRYVIGVDAFNWSKWSRESLENTIA 171  
QY 193 DSK--DHVPVNDGFOALOGIITHSKSCIELLAEPSTICAGESFQVVRNGNFRHA 249  
Db 172 SKPCGDHVFQVNN-FAVKTIONLQEKI-----FAIEGTGTGSTSFCEMSQEGFSAA 225

RESULT 13  
Q9UA13

ID Q9UA13 PRELIMINARY; PRT; 3767 AA.  
AC Q9UA13; Q21340;  
DT 01-MAY-2000 (T-EMBLrel. 13, Created)  
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
DE TRANSMEMBRANE CELL ADHESION RECEPTOR MUA-3 PRECURSOR (K08E5.3 PROTEIN)  
DE (FRAGMENT).  
DN MUA-3 OR K08E5.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Lu Z., Vogel B., Hedgecock E.;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Kershaw J.K.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
RT investigating biology";  
RL Science 282:2012-2018(1998).  
DR EMBL: AF139060; AAD29428.1; -.  
DR EMBL: Z30974; CA83226.2; -.  
DR EMBL: Z30423; CA83226.2; JOINED.  
DR HSSP: P01130; LDL.  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR001762; Disintegrin.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR001881; EGF\_Ca.  
DR InterPro: IPR002172; LDL\_recept\_A.  
DR InterPro: IPR000082; SEA.  
DR Pfam: PF00008; EGF; 33.  
DR Pfam: PF00057; ldl\_recept\_a; 4.  
DR Pfam: PF01390; SEA; 2.  
DR Pfam: PF00092; vwa; 1.  
DR PRINTS: PR00289; DISINTEGRIN.  
DR PRINTS: PR00261; LDLRECEPTOR.  
DR PRINTS: PR00453; VWFADOMAIN.  
DR SMART: SM00181; EGF; 51.  
DR SMART: SM00179; EGF\_Ca; 1.  
DR SMART: SM00001; EGF\_like; 45.  
DR SMART: SM00192; LDLa; 4.  
DR SMART: SM00200; SEA; 2.  
DR SMART: SM00327; VWA; 1.  
DR PROSITE: PS00010; ASX\_HYDROXYL; UNKNOWN\_32.  
DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.  
DR PROSITE: PS01186; EGF\_2; 6.  
DR PROSITE: PS01187; EGF\_Ca; 1.  
DR PROSITE: PS50068; LDLRA\_2; 2.  
DR PROSITE: PS50024; SEA; 4.  
DR PROSITE: PS50234; VWFA; 2.  
KW Calcium-binding; EGF-like domain; Glycoprotein; Receptor; Repeat;  
KW Signal.  
FT SIGNAL 1 24 POTENTIAL.  
FT CHAIN 25 >3767 3.  
FT NON\_TER 3767 3767  
FT SEQUENCE 3767 AA; 417284 MW; 8DA3AE5A50AEB8E CRC64;

Query Match 7.4%; Score 128; DB 5; Length 3767;  
Best Local Similarity 27.1%; Pred. No. 0.16; Indels 66; Gaps 13;  
Matches 64; Conservative 31; Mismatches 75; Indels 66; Gaps 13;





us-09-970-076-8.rspt

Fri Aug 9 10:57:09 2002

-----

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:34:18 ; Search time 61.73 Seconds  
(without alignments)  
518.350 Million cell updates/sec

Title: US-09-970-076-8  
Perfect score: 1728  
Sequence: 1 MATAERRALGIGFQWLSLAT.....TTHCSLHKIASGPTTAACME 333

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71.\*

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150.5	8.7	724	2 A48569	antigen Em100 - Ei
2	145.5	8.4	1153	1 RWHU1B	cell surface glyco
3	143.5	8.3	1163	1 RWHU1C	cell surface glyco
4	143	8.3	3124	1 A40020	collagen alpha 1(X
5	141	8.2	1153	2 S00551	leukocyte surface
6	137.5	8.0	712	2 A45638	immunodominant mic
7	137	7.9	1747	1 A45974	collagen alpha 1(X
8	131	7.6	760	1 C2MS	classical-compleme
9	130	7.5	1857	2 S31212	collagen alpha 1(X
10	130	7.5	1888	2 S78476	collagen alpha 1(X
11	128	7.4	3051	2 S42373	hypothetical prote
12	120.5	7.0	496	2 A37979	cartilage matrix p
13	118	6.8	929	2 I51027	type XII collagen
14	113.5	6.6	764	1 BBHU	complement factor
15	111.5	6.5	2944	2 A54849	collagen alpha 1(V
16	110.5	6.4	761	1 BMS	complement factor
17	109.5	6.3	1029	1 S21369	collagen alpha 2(V
18	108.5	6.3	500	2 S66522	cartilage matrix p
19	107.5	6.2	493	2 A33809	cartilage matrix p
20	106.5	6.2	747	2 I51579	complement factor
21	105.5	6.1	567	2 T28787	hypothetical prote
22	100.5	5.8	574	2 A46283	sporozoite surface
23	100	5.8	460	2 T23087	hypothetical prote
24	100	5.8	537	2 T04822	hypothetical prote
25	99.5	5.8	340	2 E70121	hypothetical prote
26	99.5	5.8	597	2 S33578	rop protein - frui
27	99	5.7	559	2 S04531	thrombospondin-rel
28	99	5.7	698	2 B96958	dnak protein (heat
29	99	5.7	741	2 T46488	hypothetical prote

## ALIGNMENTS

RESULT 1

A48569  
antigen Em100 - Eimeria maxima  
C:Species: Eimeria maxima  
C:Date: 01-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Mar-1998  
C:Accession: A48569  
R:Pasamontes, L.; Hug, D.; Humbelin, M.; Weber, G.  
Mol. Biochem. Parasitol. 57, 171-174, 1993  
A:Title: Sequence of a major Eimeria maxima antigen homologous to the Eimeria tenella  
A:Reference number: A48569; MUID:93149203  
A:Accession: A48569  
A:Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-724 <PAS>  
A:Cross-references: GB:M99058; NID:gl58890; PID:gl58891  
A:Note: Sequence extracted from NCBI backbone (NCBI:123776, NCBI:123777)  
C:Superfamily: thrombospondin type 1 repeat homology; von Willebrand factor type A re  
F:45-218/Domain: von Willebrand factor type A repeat homology <VWA2>  
F:238-296/Domain: thrombospondin type 1 repeat homology <THR1>  
F:309-371/Domain: thrombospondin type 1 repeat homology <THR2>  
F:372-432/Domain: thrombospondin type 1 repeat homology <THR3>  
F:433-493/Domain: thrombospondin type 1 repeat homology <THR4>  
F:494-556/Domain: thrombospondin type 1 repeat homology <THR5>  
F:560-610/Domain: thrombospondin type 1 repeat homology <THR6>

Query Match 8.7%; Score 150.5; DB 2; Length 724;

Best Local Similarity 25.9%; Pred. No. 0.00032;

Matches 69; Conservative 45; Mismatches 109; Indels 43; Gaps 17;

Qy 1 MATAERRALGIGFQWLSLATLVLCAG-----QGRREDGGPACYGDFLIDKSGS 54  
Db 1 MALLPQRLAPG--W-ALSLVFLAAGLTFHSSHAASEADOVCTRLDLVLMVVDSEGS 57  
Qy 55 V-LHWNNEIYFVEQLAHKF-ISP-QLRMSFIVFSTRTTLMKLTEDREQIRQGLELQK 111  
Db 58 ICTSYGKVRSPISNFACTMPLSPDDVRVGLVTFCTSAVTRWDLSDSRAQNADLLAAAK 117  
Qy 112 VLP--GGDTYMHGFERASEQIYYENRQYR--TASVIALFDGELHEDLFYSERE---- 164  
Db 118 KLPYAAGSTYTHGLAKA--EELFSFGKGRDNAPKMLVMTDGA-----SSRRSOT 168  
Qy 165 ---ANRSDLGAIYVCVG-VDNETQLARIA--DSKDHV-FP--VNDGFOALOGIIHSI 215  
Db 169 LSAEKLNRNGVIIVLVGTGVNSAECSRAGCDTSDFECPRYLQSNWGVSGOINGI 228  
Qy 216 LKKSCEI----LAAEPSTI--CAGE 235  
Db 229 IKAACKDLAKDVCSEWSEYGPCEGE 254

RESULT 2

Fri Aug 9 10:57:08 2002

Blood 79, 865-870, 1992  
A:Title: Characterization of the myeloid-specific CD11b promoter.  
A:Reference number: 152567; MUID:92144986  
A:Accession: 152567  
A:Status: translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-9 <RES>  
A:Cross-references: GB:M84477; NID:gl80184; PIDN:AAA51960.1; PID:g553219  
C:Comment: A common beta chain (CD18) forms a heterodimer with this chain to form Mac  
C:Genetics: OMIM:120980  
A:Gene: GDB:ITGAM; CR3A  
A:Cross-references: GDB:120599; OMIM:120980  
A:Map position: 16p11.2-16p11.2  
A:Note: promoter contains a GATA motif and two Sp1 consensus binding sites  
C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat h  
C:Keywords: alternative splicing; calcium; cell adhesion; glycoprotein; heterodimer;  
F:1-16/Domain: signal sequence #status predicted <EXT>  
F:17-1108/Domain: extracellular #status predicted <EXT>  
F:148-318/Domain: von Willebrand factor type A repeat homology <VWA2>  
F:465-473/Region: calcium/magnesium binding #status predicted  
F:530-538/Region: calcium/magnesium binding #status predicted  
F:593-601/Region: calcium/magnesium binding #status predicted  
F:1109-1134/Domain: transmembrane #status predicted <TM>  
F:1135-1153/Domain: intracellular #status predicted <INT>  
F:86,240,391,469,693,697,735,802,881,901,912,941,947,979,994,1022,1045,1051,1076/Bind

Query Match 8.4%; Score 145.5; DB 1; Length 1153;  
Best Local Similarity 26.2%; Pred. No. 0.0015;  
Matches 60; Conservative 45; Mismatches 81; Indels 43; Gaps 12;

QY 44 DLYFILDKSSVIL-HHWNIEYFVLEQLAHKFIHQKMSFIVFSTGTMLKLTED---- 98  
DB 150 DIAFLIDGSGSIIPHDFRMKEVST-----VMEQLKSKTLFS-----LMQYSEEFRIH 199  
QY 99 -----REQIRGLEELQKVLPGDGTVMHGFERASQIYYENRQGYRTA-SVITALT 149  
DB 200 FTFKEFQNNPRLSKVPITQLL--GRHTATGIRKVVRELFNITGARKNAFKILVVIT 257  
QY 150 DGEHEDLFFYSE--REANRSRLGAIYCVGVKDFNETQLAR-----IADS--KDHVPF 200  
DB 258 DGEKFGDPLGYEDVPEADRE--GVIRYVIGVGDAFRSEKSRQELNTIASKPPRDHFVQ 314  
QY 201 VNDGFOALQGIHHSILKSKICIELAEPSTICAGESFOVVVRGNGFRHA 249  
DB 315 VNN-FAELKTIONLREK----IFAIECTGTGSSSPFHEMSQEGFSAA 358

RESULT 3  
RWUHC  
cell surface glycoprotein CD11c precursor - human  
N:Alternate names: leukocyte adhesion receptor p150,95 alpha chain  
C:Species: Homo sapiens (man)  
C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 22-Jun-1999  
C:Accession: A36584; A35543; S00864  
R:Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.  
J. Biol. Chem. 265, 12750-12751, 1990  
A:Reference number: A36584  
A:Accession: A36584  
A:Contents: erratum  
A:Accession: A36584  
A:Molecule type: DNA  
A:Residues: 1-1163 <COR>  
A:Note: this revision to the sequence from reference A35543 includes the carboxyl end  
R:Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.  
J. Biol. Chem. 265, 2782-2786, 1990  
A:Title: Genomic structure of an integrin alpha subunit, the leukocyte p150,95 molecu  
A:Reference number: A35543; MUID:90153906  
A:Accession: A35543  
A:Molecule type: DNA  
A:Residues: 1-834 <CO2>  
A:Note: this sequence has been revised in reference A36584  
R:Corbi, A.L.; Miller, L.J.; O'Connor, K.; Larson, R.S.; Springer, T.A.

RWUHC  
cell surface glycoprotein CD11b precursor [validated] - human  
N:Alternate names: complement receptor type 3 alpha chain; leukocyte adhesion protein Ma  
eukocyte integrin alpha chain; neutrophil adherence receptor alpha chain  
C:Species: Homo sapiens (man)  
C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 31-Dec-2000  
C:Accession: A31108; A28915; A41600; A30892; A32218; A46526; A26091; 152567  
R:Corbi, A.L.; Kishimoto, T.K.; Miller, L.J.; Springer, T.A.  
J. Biol. Chem. 263, 12403-12411, 1988  
A:Title: The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD  
B.  
A:Reference number: A31108; MUID:88315033  
A:Accession: A31108  
A:Molecule type: mRNA  
A:Residues: 1-1153 <COR>  
A:Cross-references: GB:J03925; NID:gl87284; PIDN:AAA59544.1; PID:g307148  
A:Note: part of this sequence was confirmed by protein sequencing  
R:Arnaout, M.A.; Gupta, S.K.; Pierce, M.W.; Tenen, D.G.  
J. Cell Biol. 106, 2153-2158, 1988  
A:Title: Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor M  
A:Reference number: A28915; MUID:88257215  
A:Accession: A28915  
A:Molecule type: mRNA  
A:Residues: 1-499,501-965, P', 967-1153 <ARN>  
A:Cross-references: GB:M19664; GB:J03270; NID:gl86935; PIDN:AAA594  
A:Note: the authors translated the codon TAC for residue 1129 as Thr  
R:Arnaout, M.A.; Remold-O'Donnell, E.; Pierce, M.W.; Tenen, D.G.  
Proc. Natl. Acad. Sci. U.S.A. 85, 2776-2780, 1988  
A:Title: Molecular cloning of the alpha-subunit of human and guinea pig leukocyte adhesi  
A:Reference number: A94193; MUID:88190151  
A:Accession: A30892  
A:Molecule type: mRNA  
A:Residues: 917-1042 <AR2>  
A:Cross-references: GB:M76724; NID:gl80018; PIDN:AAA58410.1; PID:g553215  
R:Arnaout, M.A.; Remold-O'Donnell, E.; Pierce, M.W.; Tenen, D.G.  
Proc. Natl. Acad. Sci. U.S.A. 85, 2776-2780, 1988  
A:Title: cDNA sequence for the alpha subunit of the human neutrophil adherence receptor  
A:Reference number: A32218; MUID:89098893  
A:Accession: A32218  
A:Molecule type: mRNA  
A:Residues: 9-1153 <HIC>  
A:Cross-references: GB:J04145; NID:gl89068; PIDN:AAA59903.1; PID:g386975  
A:Note: part of this sequence was confirmed by protein sequencing  
R:Fleming, J.C.; Pahl, H.L.; Gonzalez, D.A.; Smith, T.F.; Tenen, D.G.  
J. Immunol. 150, 480-490, 1993  
A:Title: Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-in  
n during evolution.  
A:Reference number: A46526; MUID:93123748  
A:Accession: A46526  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-499,501-1153 <FLE>  
A:Cross-references: GB:S52227; NID:g263047; PIDN:AAB24821.1; PID:g263049  
A:Note: the last three bases of intron 13, CAG, are included in some but not all mature  
R:Pierce, M.W.; Remold-O'Donnell, E.; Todd III, R.F.; Arnaout, M.A.  
Blood. Biophys. Acta 874, 368-371, 1986  
A:Title: N-terminal sequence of human leukocyte glycoprotein Mol: conservation across sp  
A:Reference number: A90664; MUID:87076671  
A:Accession: A26091  
A:Molecule type: protein  
A:Residues: 17-31 <PIE>  
A:Experimental source: granulocytes  
R:Pahl, H.L.; Rosmarin, A.G.; Tenen, D.G.





R;Gerecke, D.R.; Foley, J.W.; Castagnola, P.; Gennari, M.; Dublet, B.; Cancedda, R.; Llin  
J. Biol. Chem. 268, 12177-12184, 1993  
A:Title: Type XIV collagen is encoded by alternative transcripts with distinct 5' region  
ns.

A:Reference number: A45974; MUID:93280195  
A:Accession: A45974  
A:Status: preliminary  
A:Molecule type: mRNA; protein  
A:Residues: 1-1747 <GER>  
A:Experimental source: embryo skin  
A>Note: sequence inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBI:133364, NCBIP:133365)  
R;Apte, S.S.  
submitted to the EMBL Data Library, March 1992

A:Reference number: S30085  
A:Accession: S30085  
A:Molecule type: mRNA  
A:Residues: 1472-1660 <APT>  
A:Cross-references: EMBL:X65122; NID:g62871; PIDN:CAA46238.1; PID:g938175  
R;Trueb, J.; Trueb, B.

Eur. J. Biochem. 207, 549-557, 1992  
A:Title: Type XIV collagen is a variant of undulin.  
A:Reference number: S22916; MUID:92339443  
A:Accession: S22916  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 286-494, 'Q', 496-834, 'A', 836-1119, 'KL', 1122-1402, 1409-1439 <TRU>  
R;Gordon, M.K.; Castagnola, P.; Dublet, B.; Linsemayer, T.F.; van der Rest, M.; Mayne,  
Eur. J. Biochem. 201, 333-338, 1991

A:Title: Cloning of a cDNA for a new member of the class of fibril-associated collagens  
A:Reference number: S17035; MUID:92037585

A:Accession: S17035  
A:Molecule type: mRNA  
A:Residues: 1472-1659 <GORI>  
A:Accession: S20833  
A:Molecule type: protein  
A:Residues: 1551-1570; 1593-1667 <GOR>  
C:Superfamily: collagen alpha 1(XIV) chain; fibronectin type III repeat homology; von Willebrand factor type A repeat homology; triple helical domain  
C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trim  
F:40-204/Domain: von Willebrand factor type A repeat homology <VWAI>  
F:236-317/Domain: fibronectin type III repeat homology <FN3A>  
F:326-409/Domain: fibronectin type III repeat homology <FN3B>  
F:418-498/Domain: fibronectin type III repeat homology <FN3C>  
F:507-591/Domain: fibronectin type III repeat homology <FN3C>  
F:625-707/Domain: fibronectin type III repeat homology <FN3B>  
F:716-798/Domain: fibronectin type III repeat homology <FN3B>  
F:806-893/Domain: fibronectin type III repeat homology <FN3B>  
F:924-1089/Domain: von Willebrand factor type A repeat homology <VWA2>  
F:1111-1352/Domain: non-collagenous NC4 #status predicted <NC4>  
F:1511-1553/Domain: non-collagenous NC2 #status predicted <NC2>  
F:1554-1659/Domain: triple helical domain COL1 #status predicted <COL1>

Query Match 7.9%; Score 137; DB 1; Length 1747;  
Best Local Similarity 25.4%; Pred. No. 0.013;  
Matches 63; Conservative 45; Mismatches 114; Indels 26; Gaps 12;

QY 44 DLVFLDKSGV-LHHWNEIYFVEQL--AHKFTSPQ-LRMSFVSTRTTLMKLT--E 97  
DB 926 DLVFLVDCGWSIGDDNFNKNISLYSTVFGALDKIGDGTQVATIQFSDPRTEFKLNAYK 985  
QY 98 DREIROGLELEOKVLPGGDTYMHGPERASEQIYYENROGYRTA--SVIIALTQDGLHE 155  
DB 986 TKETLEIAIQI--AYKGGNTKTKAIKHARE-VLFTGEAGMRKGPVKLVITDGRSQD 1042  
QY 156 DLFFYSREANRSDL-GAIVYCVGVKDFNETQLARIAD--SKDHVPPVNDGFQALQGII 212  
DB 1043 DV-----NKVSREMQLDCGFFAIGVADADYSELVNIKSPKPSERHVFVDD-EDAFTKIE 1096  
QY 213 HSILKKSCEILAAEPTICAGESFQVVRGNGFRHARNVDRVLCSPKINDSVTLNEKPF 272  
DB 1097 DELITFVCETASATCPLVFLFKDGNFA-----GFKMMEMFGLVEKFSADIDGVSMEPGTF 1150

QY 273 SVEDTYLL 280  
DB 1151 NVYPCYRL 1158

## RESULT 8

## C2MS

classical-complement-pathway C3/C5 convertase (EC 3.4.21.43) C2 component precursor -  
N:Alternate names: C3 convertase; C5 convertase; complement C2  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 19-May-2000  
C:Accession: A38876; B36593; I54429  
R;Ishikawa, N.; Nonaka, M.; Wetsel, R.A.; Colten, H.R.  
submitted to GenBank, January 1991  
A:Reference number: A38875  
A:Accession: A38876  
A:Molecule type: DNA  
A:Residues: 1-760 <IS2>  
A:Cross-references: GB:M57891; GB:J05661; NID:g192436; PIDN:AAA63294.1; PID:g192437  
R;Ishikawa, N.; Nonaka, M.; Wetsel, R.A.; Colten, H.R.  
J. Biol. Chem. 265, 19040-19046, 1990  
A:Title: Murine complement C2 and factor B genomic and cDNA cloning reveals different  
A:Reference number: A36593; MUID:91035430  
A:Accession: B36593  
A:Molecule type: mRNA  
A:Residues: 1-760 <ISH>  
A:Cross-references: EMBL:M57891; NID:g192436; PIDN:AAA63294.1; PID:g192437  
R;Falus, A.; Wakeland, E.K.; McConnell, T.J.; Gitlin, J.; Whitehead, A.S.; Colten, H.  
Immunogenetics 25, 290-298, 1987  
A:Title: DNA polymorphism of MHC III genes in inbred and wild mouse strains.  
A:Reference number: I54429; MUID:87192938  
A:Accession: I54429  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 660-677, 'R', 679, 681-723, 'G', 725 <RES>  
A:Cross-references: GB:M16271; NID:g199289; PIDN:AAA9562.1; PID:g199290  
C:Genetics:  
A:Introns: 16/1; 91/1; 153/1; 212/1; 245/3; 290/3; 337/1; 384/1; 461/1; 492/3;  
C:Complex: The proenzyme forms a complex with C4a and is activated by cleavage into C  
C:Function:  
A:Description: cleaves complement C3 and complement C5 alpha chains  
A:Pathway: complement classical pathway  
C:Superfamily: complement C2; complement factor H repeat homology; trypsin homology;  
C:Keywords: alternative splicing; complement classical pathway; duplication; glycopro  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-250/Product: complement C2b fragment #status predicted <C2B>  
F:22-89/Domain: complement factor H repeat homology <FHL>  
F:94-149/Domain: complement factor H repeat homology <FHL>  
F:156-210/Domain: complement factor H repeat homology <FHL>  
F:251-760/Product: complement C2a fragment long form #status predicted <C2A>  
F:251-605, 613-760/Product: complement C2a fragment short form #status predicted <C2S>  
F:259-449/Domain: von Willebrand factor type A repeat homology <VFA>  
F:478-747/Domain: trypsin homology #status atypical <TRY>  
F:22-62, 49-89, 94-136, 122-149, 156-197, 182-210, 470-590, 499-515, 593-609, 647-674, 685-715/  
F:27, 117, 297, 340, 474, 478, 663/Binding site: carbohydrate (Asn) (covalent) #status pred  
F:514, 570, 689/Active site: His, Asp, Ser #status predicted

Query Match 7.6%; Score 131; DB 1; Length 760;  
Best Local Similarity 23.2%; Pred. No. 0.013;  
Matches 74; Conservative 56; Mismatches 115; Indels 74; Gaps 16;

QY 41 GGFDPFLDKSGV-LHHWNEIY-YFVEQLAHKFTSPQ-LRMSFVSTRTTLMKLT 97  
DB 258 GHUNLYLLDASQSVTEKDFDIFKKSALMVERIFSEVNVVAITTFASOPKTIMILS 317  
QY 98 DREIROGLELEOKVLPGGDTYMHGPERASEQIYYE-----NROGYRTAS- 143  
DB 318 ERSQ-----DVTEVITSLDSASVKDHENATGANTYELVIRVYSMMQTQMDRLGMEISAW 371  
QY 144 -----VVIATDGLHEDLFFYSREANRSDLGAI-----VYCVGV-----KD 182  
DB 372 KEIRHTIILLTDGK--SNMGDSPKKAIVTRIRELLSIEQNRRDDYLDIYAIGVKLDVDWKE 429





A:Reference number: S42368

A:Accession: S42373

A:Molecule type: DNA

A:Residues: 1-3051 <SM>

A:Cross-references: EMBL:Z30423; NID:g458479; PID:g458485

C:Genetics:

C:Introns: 96/1; 166/3; 201/2; 355/1; 404/1; 427/2; 794/1; 892/1; 952/2; 1041/1; 1146/1;

C:Superfamily: von Willebrand factor type A repeat homology; EGF homology; fibronectin

F:512-679/Domain: von Willebrand factor type A repeat homology <vwa1>

F:754-793/Domain: fibronectin type II repeat homology <2f1>

F:1201-1244/Domain: EGF homology <EGF>

Query Match 7.4%; Score 128; DB 2; Length 3051;

Best Local Similarity 27.1%; Pred. No. 0.15;

Matches 64; Conservative 31; Mismatches 75; Indels 66; Gaps 13;

QY 39 CYGGF-----DLYFLDKSGSVLHH--WNEIYFYFEQLA 70

Db 483 CYGGFVDSSNANLPGRVCTVQTCPKQKTDLVFLIDGSGSIGSVFKNEVLRFEV 542

QY 71 HKFI--SPQLRMSFVIFSTRTLMKLT--DREQIRGLEELQKVLPG-----GDYIMH- 121

Db 543 ELFEIGRSRTRVGLIQSDQIRHEFDLQDGRDRLKGISETQ-YLTGLTRTGAIAQHM 601

QY 122 --EGFERASEQIYYENKQYR-----TASVIALTGELHEDLFFYSEREANRDLGAI 174

Db 602 VQEGF-----SRRGARPOSDIARVAIILTDGRSQDNV-----TGPADSARKLSIN 648

QY 175 VYCVGVKVD-FNEPQLARIADSKDHVPVNDGFGFQALGGIHSILKSKSCIEILAAEPS 229

Db 649 TFAIGVTDHVLASELSIAGSPNRWFYV-DKPKDLDLRLSRMIQK-----AACPS 697

RESULT 12

A37979

cartilage matrix protein precursor - human

C:Species: Homo sapiens (man)

C:Date: 12-Jul-1991 #sequence\_revision 12-Jul-1991 #text\_change 21-Jul-2000

C:Accession: A37979; B37979

R.Jenkins, R.N.; Osborne-Lawrence, S.L.; Sinclair, A.K.; Eddy Jr., R.L.; Byers, M.G.; Sh

J. Biol. Chem. 265, 19624-19631, 1990

A:Title: Structure and chromosomal location of the human gene encoding cartilage matrix

A:Reference number: A37979; MUID:91060568

A:Accession: A37979

A:Molecule type: DNA

A:Residues: 1-496 <JEN>

A:Cross-references: GB:J05667

A:Accession: B37979

A:Molecule type: mRNA

A:Residues: 157-290, 'L', 292-496 <JE2>

A:Cross-references: GB:M55683; GB:J05666; GB:J05667; NID:g180651; PID:g

C:Genetics:

A:Gene: GDB:CRTM

A:Cross-references: GDB:127280; OMIM:115437

A:Map position: 1p35-1p35

A:Introns: 32/1; 147/3; 222/1; 264/1; 403/1; 454/1; 481/1

A:Complex: homotrimer

C:Superfamily: unassigned EGF-related proteins; EGF homology; von Willebrand factor type

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-496/Product: cartilage matrix protein #status predicted <MAT>

F:39-206/Domain: von Willebrand factor type A repeat homology <vwa1>

F:227-262/Domain: EGF homology <EGF>

F:273-437/Domain: von Willebrand factor type A repeat homology <vwa2>

F:76,344/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:221-238,234-247,249-262/Disulfide bonds: #status predicted

Query Match

Best Local Similarity 25.3%; Pred. No. 0.055;

Matches 60; Conservative 37; Mismatches 95; Indels 45; Gaps 13;

QY 24 ICAGQGRREDGPGACYGGFDLYFLDKSGSVL-HHNEIYFYFEQLAHKF-ISPQLRMS 81

Db 264 VCSGGGS-----SATDLVFLIDGSKSVRPENFELVKKFISIVDTLDVSDKLAQV 314

QY 82 FIV---FSTRGTTLMKLTEDRQIROGLEELQKVLPGDGTTHMHEGFERASEQIY-----Y 133

Db 315 GLVOYSSSVRQEPFLGRFHTTKDKIAAVRNM-----SYMEKGTMTGAALKYLDNSF 366

QY 134 ENRQGYRTAS--VIALTGELHEDLFFYSEREANRDLGAIYCVGVKDFNETOLARI 191

Db 367 TVSSGARPGAQKVGIVFTDGRSOD-----YINDAAKAKDLGFKMFAVGONAVEDELEI 422

QY 192 ADS--KDHVFPVNDGFGFQALGGIHSILKSKSCIEILAAEPSTICAGES---FQVYVRG 243

Db 423 ASEPVAEHYFYPAD-FKTINQIGKQLQKICVE---EDP---CACESLVKFOAKVEG 472

RESULT 13

I51027

type XII collagen alpha-1 chain - eastern newt (fragment)

C:Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)

C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jan-2000

C:Accession: I51027

R.Wei, Y.; Yang, E.V.; Klatt, K.P.; Tassava, R.A.

Dev. Biol. 168, 503-513, 1995

A:Title: Monoclonal antibody MT2 identifies the urodele alpha 1 chain of type XII col

A:Reference number: I51027; MUID:95246925

A:Accession: I51027

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-929 <WEI>

A:Cross-references: EMBL:U19494; NID:g632647; PID:AAA80217.1; PID:g632648

C:Superfamily: collagen alpha 1(XII) chain; fibronectin type III repeat homology; von

F:155-236/Domain: fibronectin type III repeat homology <3FR>

F:631-795/Domain: von Willebrand factor type A repeat homology <vwa3>

Query Match 6.8%; Score 118; DB 2; Length 929;

Best Local Similarity 21.2%; Pred. No. 0.2;

Matches 57; Conservative 51; Mismatches 135; Indels 26; Gaps 9;

QY 44 DLYFLDKSGSV-LHHNEIYFYFEQLAHKFISPQLRMSFVIFSTRTLMKLTEDREQI 102

Db 633 DLYLLVDGWSIGRPNFKIVRNFISRVVEFDIGSDRVQIAVSQYSGDPRTTEWQLNTHKT 692

QY 103 RQGLEELQKVLPG--GGDTYMHGFERASEQIYYENRQGY-----RTASVIALTGELHED 156

Db 693 KKSMDAVANLPYKGGTNTGTSALKFILENNF---RPGVGMREKARKIAILLTDGKSQDD 749

QY 157 LFFYSEREANRDLGAIYCVGVKDFNETQLARIADSKDHVPVN-DGFOALQGIHISI 215

Db 750 IVAPSKRYA----DEGIELYAVGIKKNADENELKEIASDPDELYMYNVADESLLTINVDL 805

QY 216 LKKSCEILAA-----EPSTICAGESFQVYVRGNGFRHARNVDRVLCFSFKINDSVTLNEKP 271

Db 806 TENVCNSVKGPGGLNPNPSNLVTSEPTSPRSPRTVWPPSQSVR----FKVEIYYPVAGGRP 861

QY 272 FSVEDYIYLLCPAPILKEVGNKAAQLQVSMN 300

Db 862 ---QEVYVRGTQTTLVLVLGLKPKETEVYVN 887

RESULT 14

BBHU

complement factor B precursor [validated] - human

N:Alternate names: C3 convertase; C3 proactivator; glycine-rich beta-glycoprotein; he

N:Contains: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) Bb fragmen

C:Species: Homo sapiens (man)

C:Date: 19-Feb-1984 #sequence\_revision 05-Aug-1994 #text\_change 08-Dec-2000

C:Accession: S34075; A44622; A00934; A19188; A19947; B19947; B25971; S14339; A44628;

R:Mejta, J.E.; Jahn, I.; de la Salle, H.; Hauptmann, G.

submitted to the EMBL Data Library, March 1993

A:Reference number: S34075

A:Accession: S34075  
A:Molecule type: mRNA  
A:Residues: 1-764 <MEJ>  
A:Cross-references: EMBL:X72875; NID:g297568; PIDN:CAA51389.1; PID:g297569  
R:Woods, D.E.; Markham, A.F.; Ricker, A.T.; Goldberger, G.; Colten, H.R.  
Proc Natl. Acad. Sci. U.S.A. 79, 5661-5665, 1982  
A:Title: Isolation of cDNA clones for the human complement protein factor B, a class III  
A:Reference number: A44622; MUID:83039428  
A:Accession: A44622  
A:Molecule type: mRNA  
A:Residues: 467-546;550-595;752-764 <WOO>  
A:Cross-references: GB:J00185; GB:J00186  
A:Note: the authors translated the codon TAC at 519 as Thr; the nucleic acid translation  
R:Mole, J.E.; Anderson, J.K.; Davison, E.A.; Woods, D.E.  
J. Biol. Chem. 259, 3407-3412, 1984  
A:Title: Complete primary structure for the zymogen of human complement factor B.  
A:Reference number: A20751; MUID:84161997  
A:Accession: A00934  
A:Molecule type: protein; mRNA  
A:Residues: 26-764 <MOI>  
A:Cross-references: GB:K01566  
A:Note: nucleic acid translation differs from the sequence shown in having 300-Leu, 328-  
A:Note: 736-Ser was also found  
A:Note: glycosylation sites were determined  
R:Christie, D.L.; Gagnon, J.  
Biochem. J. 209, 61-70, 1983  
A:Title: Amino acid sequence of the Bb fragment from complement factor B. Sequence of th  
A:Reference number: A19188; MUID:83204002  
A:Contents: the final paper in a series documenting the sequence, glycosylation site, an  
A:Accession: A19188  
A:Molecule type: protein  
A:Residues: 260-296, 'T', 298-764 <CHR>  
R:Campbell, R.D.; Porter, R.R.  
Proc. Natl. Acad. Sci. U.S.A. 80, 4464-4468, 1983  
A:Title: Molecular cloning and characterization of the gene coding for human complement  
A:Reference number: A19947; MUID:83273641  
A:Accession: A19947  
A:Molecule type: DNA  
A:Residues: 346-764 <CAM>  
A:Cross-references: GB:J00125  
A:Accession: B19947  
A:Molecule type: mRNA  
A:Residues: 339-509 <CAI>  
A:Cross-references: GB:J00126; NID:g187723; PIDN:AAA36226.1; PID:g553536  
R:Wu, L.; Morley, B.J.; Campbell, R.D.  
Cell 48, 331-342, 1987  
A:Title: Cell-specific expression of the human complement protein factor B gene: eviden  
A:Reference number: A25971; MUID:87102880  
A:Accession: B25971  
A:Molecule type: DNA  
A:Residues: 1-99 <WUL>  
A:Cross-references: GB:M15082; NID:g187699; PIDN:AAA59625.1; PID:g553534  
R:Niemann, M.A.; Brown, A.S.; Miller, E.J.  
Biochem. J. 274, 473-480, 1991  
A:Title: The principal site of glycation of human complement Factor B.  
A:Reference number: S14339; MUID:91174758  
A:Accession: S14339  
A:Molecule type: protein  
A:Residues: 270-329 <NIE>  
A:Note: binding site for carbohydrate to lysine under artificial conditions  
R:Morley, B.J.; Campbell, R.D.  
EMBO J. 3, 153-157, 1984  
A:Title: Internal homologies of the Ba fragment from human complement component factor B  
A:Reference number: A44628; MUID:84158524  
A:Accession: A44628  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 16-225, 'F', 227-259 <WOR>  
R:Schwaible, W.; Lüttig, B.; Sokolowski, T.; Estaller, C.; Weiss, E.H.; Meyer zum Busche  
Immunobiology 188, 221-232, 1993  
A:Title: Human complement factor B: functional properties of a recombinant zymogen of th  
A:Reference number: I54409; MUID:94041399  
A:Accession: I54409

A>Status: translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-764 <RES>  
A:Cross-references: GB:S67310; NID:g452937; PIDN:AAI3989.1; PID:g4261689  
R:Horiuchi, T.; Kim, S.; Matsumoto, M.; Watanabe, I.; Fujita, S.; Volanakis, J.E.  
Mol. Immunol. 30, 1587-1592, 1993  
A:Title: Human complement factor B: cDNA cloning, nucleotide sequencing, phenotypic c  
A:Reference number: I57824; MUID:94067177  
A:Accession: I57824  
A>Status: translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-31, 'Q', 33-764 <RE2>  
A:Cross-references: GB:L15702; NID:g291921; PIDN:AAA16820.1; PID:g291922  
C:Comment: 292-Cys has a free sulphydryl.  
C:Genetics:  
A:Gene: GDB:BF  
A:Cross-references: GDB:I19726; OMIM:138470  
A:Map position: 6p21.3-6p21.3  
A:Introns: 21/3; 99/3; 346/1; 470/1; 424/1; 390/1; 502/3; 542/1; 593/2; 619/1; 652/3;  
A:Note: the list of introns may be incomplete  
A:Note: gene is located in the major histocompatibility complex, class III region  
C:Complex: complement factor B initially forms an inactive complex with complement fa  
ment factor C3b forming active C3/C5 convertase; Ba is released  
C:Function:  
A:Description: Bb is a serine proteinase; C3/C5 convertase cleaves complement C3 alph  
al  
A:Pathway: complement alternate pathway  
C:Superfamily: complement C2; complement factor H repeat homology; trypsin homology;  
C:Keywords: acute phase; complement alternate pathway; duplication; glycoprotein; hyd  
F:1-25/Domain: signal sequence #status predicted <SIG>  
F:26-764/Product: complement factor B #status experimental <MAT>  
F:26-259/Product: complement factor Ba fragment #status experimental <BAF>  
F:37-98/Domain: complement factor H repeat homology <FH1>  
F:103-158/Domain: complement factor H repeat homology <FH2>  
F:165-218/Domain: complement factor H repeat homology <HH>  
F:260-764/Product: C3/C5 convertase Bb fragment #status experimental <VFA>  
F:268-458/Domain: von Willebrand factor type A repeat homology <FHD>  
F:482-752/Domain: trypsin homology #status atypical <TRY>  
F:37-76, 62-98, 103-145, 131-158, 165-205, 191-218, 478-596, 511-527, 599-615, 656-682, 695-725  
F:122, 142, 285, 378/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:259-260/Cleavage site: Arg-Lys (complement factor D) #status experimental  
F:526, 576, 699/Active site: His, Asp, Ser #status experimental

Query Match	6.6%	Score 113.5;	DB 1;	Length 764;
Best Local Similarity	19.6%	Pred. NO. 0.37;		
Matches	66;	Conservative	63;	Mismatches 109; Indels 99; Gaps 17;
QY	28	QGGRRDGG-GPA-----	CYGGFDLYFILDKSGSV-----	LHHWNEIYFFVEQLAH 71
DB	243	EGVDAEDGHGPGGQQRKIVLPDPSGMNIIYLVLDGSDSIGASNFTCAKKCLVNLIEKVAS	302	
QY	72	KFTSPQLRMSFIVFSTRGTTLMKLT-----	DREIQRGLLEEL-----	OKVLPQGDYIMHEG 123
DB	303	YGKVP--RYGLVTYATYPIKWVSEADSNADWVTKQLNEINYEDHKLKSGTNT-----	355	
QY	124	FERASEQIYYENR-----	QGY-RTASVILIALTDG-----	ELHEDLFFYSE 162
DB	356	--KKALQAVYSMSWPDVPPGWNRRHVVILMTDGLHNMGGDPITVIDEIRDLLYIGKD 414		
QY	163	REANRSRDILGAIYCVG--VKDENETQLARIADSKDHPFVNDNGFOALOGITHSLKKSC 220		
DB	415	RKNPRELDVYVFGVGPLVNOVNINALASKDNEQHFVKVD--MENLEDVFYQIMIDES- 472		
QY	221	IEILAAEPSTICAGESFOVVVRNGGFHARNVDRLCLSPKINDSVTLNKPSPSVEPTYLL 280		
DB	473	-----QSLSLC-----	GMVWEHRRKGTD-----	YHKQPQWAKISV-- 501
QY	281	CPAPILKEVGMKAALQVSMNDGLSFSSSSSVIITTHC 317		
DB	502	-----IRPSKGHESCMG-----	AVVSEYFVLTAHC 527	

## RESULT 15

A54849  
 collagen alpha 1(VII) chain precursor - human  
 N:Alternate names: procollagen alpha 1(VII) chain  
 C:Species: Homo sapiens (nan)  
 C:Date: 04-Nov-1994 #sequence\_revision 04-Nov-1994 #text\_change 20-Sep-1999  
 C:Accession: A54849; PH0844; S16316; I56328; A30296; I84686  
 R:Christiano, A.M.; Greenspan, D.S.; Lee, S.; Uitto, J.  
 J. Biol. Chem. 269, 20256-20262, 1994  
 A:Title: Cloning of human type VII collagen. Complete primary sequence of the alpha1(VII)  
 A:Reference number: A54849; MUID:94327588  
 A:Accession: A54849  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-2944 <CHR>  
 A:Cross-references: GB:I02870; NID:987124; PIDN:AAA75438.1; PID:9987125  
 R:Tanaka, T.; Takahashi, K.; Furukawa, F.; Imamura, S.  
 Biochem. Biophys. Res. Commun. 183, 958-963, 1992  
 A:Title: Molecular cloning and characterization of type VII collagen cDNA.  
 A:Reference number: PH0844; MUID:92231902  
 A:Accession: PH0844  
 A:Molecule type: mRNA  
 A:Residues: 'EFR',340-475,'RALSTASHSTLCWRATRWHPNCRGSHWTRACPEPCNRPASHRAARAG',524-528,'C',  
 A:Cross-references: DDBJ:D11152; DDBJ:DL3694; NID:g453698; PIDN:BA002853.1; PID:g453699  
 A:Experimental source: keratinocyte  
 R:Parente, M.G.; Chung, L.C.; Ryyanaenen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Mat  
 proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991  
 A:Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.  
 A:Reference number: S16316; MUID:91334380  
 A:Accession: S16316  
 A:Molecule type: mRNA  
 A:Residues: '815-892','E',894-1439 <PAR>  
 A:Cross-references: GB:M65158; GB:S49017; NID:g180914; PIDN:AAA96439.1; PID:g180915  
 A:Experimental source: keratinocyte  
 R:Gammon, W.R.; Abernethy, M.D.; Padilla, K.M.; Prissayanh, P.S.; Cook, M.E.; Wright, J.;  
 J. Invest. Dermatol. 99, 691-696, 1992  
 A:Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion prot  
 A:Reference number: I56328; MUID:93107742  
 A:Accession: I56328  
 A:Status: translated from GB/EMBL/DDBJ  
 A:Molecule type: mRNA  
 A:Residues: 'EFR',372-517,'DV',520-540,'W',542-1255 <RES>  
 A:Cross-references: GB:S51236; NID:g262308; PIDN:AAB24637.1; PID:g262309  
 R:Selzter, J.L.; Eisen, A.Z.; Bauer, E.A.; Morris, N.P.; Glanville, R.W.; Burgeson, R.E.  
 J. Biol. Chem. 264, 3822-3826, 1989  
 A:Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagen  
 A:Reference number: A30296; MUID:89139437  
 A:Accession: A30296  
 A:Molecule type: protein  
 A:Residues: 'A',1240-1246,'G',1248-1250,'XE',1253-1255,'Q',1257,'E',2032,'C',2034-2041;  
 A:Note: two reported peptides cannot be reliably located  
 R:Greenspan, D.S.  
 Hum. Mol. Genet. 2, 273-278, 1993  
 A:Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous  
 A:Reference number: I48103; MUID:93271985  
 A:Accession: I48103  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: mRNA  
 A:Residues: 2395-2871,'S',2873-2944 <RES>  
 A:Cross-references: GB:I06862; NID:g388713; PIDN:AAA89196.1; PID:g388714  
 R:Christiano, A.M.; Ryyanaenen, M.; Uitto, J.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994  
 A:Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --> Ser subs  
 A:Reference number: A55255; MUID:94224777  
 A:Contents: annotation  
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
 ed and subsequently O-glycosylated.  
 C:Genetics:  
 A:Gene: GDB:COL7A1; EBI: EBI1; EB  
 A:Cross-references: GDB:I28750; OMIM:120120  
 A:Map position: 3p21.3-3p21.3  
 A:Note: defects in this gene can result in dominant and recessive dystrophic epidermolys

A:Note: there are 118 introns  
 C:Complex: type VII collagen is probably a homotrimer  
 C:Function:

A:Description: structural component of extracellular polymer associated with anchoring  
 C:Superfamily: unassigned collagens; animal Kunitz-type proteinase inhibitor with anchoring  
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxyllysine; hydroxypro  
 F:1-16/Domain: signal sequence #status predicted <SIG>  
 F:17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>  
 F:17-1253/Domain: amino-terminal nonhelical #status predicted <NC1>  
 F:36-201/Domain: von Willebrand factor type A repeat homology <VWA1>  
 F:231-318/Domain: fibronectin type III repeat homology <FN1>  
 F:327-413/Domain: fibronectin type III repeat homology <FN2>  
 F:414-502/Domain: fibronectin type III repeat homology <FN3>  
 F:508-593/Domain: fibronectin type III repeat homology <FN4>  
 F:598-683/Domain: fibronectin type III repeat homology <FN5>  
 F:686-771/Domain: fibronectin type III repeat homology <FN6>  
 F:776-862/Domain: fibronectin type III repeat homology <FN7>  
 F:864-952/Domain: fibronectin type III repeat homology <FN8>  
 F:954-1045/Domain: fibronectin type III repeat homology <FN9>  
 F:1052-1219/Domain: von Willebrand factor type A repeat homology <VWA2>  
 F:1170-1172/Region: cell attachment (R-G-D) motif  
 F:1189-1253/Region: cysteine/proline-rich  
 F:1254-2783/Region: interrupted helical  
 F:1334-1336/Region: cell attachment (R-G-D) motif  
 F:2008-2010/Region: cell attachment (R-G-D) motif  
 F:2553-2555/Region: cell attachment (R-G-D) motif  
 F:2784-2944/Domain: carboxyl-terminal nonhelical #status predicted <NC2>  
 F:2876-2929/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>  
 F:337,786,1109/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:2167,2176,2185,2188,2664,2667,2673/Modified site: 4-hydroxyproline (Pro) #status ex  
 F:2625,2631/Binding site: 5-hydroxylysine (Lys) #status experimental  
 F:2625,2631/Binding site: carbohydrate (Lys) (covalent) #status experimental  
 F:2634,2802,2804/Disulfide bonds: interchain #status predicted

Query Match 6.5%; Score 111.5; DB 2; Length 2944;  
 Best Local Similarity 24.1%; Pred. No. 3.3;  
 Matches 52; Conservative 42; Mismatches 85; Indels 37; Gaps 11;

QY	44	DLVFILDKSGSV-LHWNHETIYFEVQLAHKF---	ISPO-LRMSFIVFSTRGTT---	LMKL	95
DB	38	DIVFLIDGSSIGRSNFRVRSFLEGLVLVPFSGAASAGQVRFATVQISDDPRTEFLDAL	97		
QY	96	TEDREQIQCLELEQLKVLPGGDYTMHEGFERASEQIYYENRQGYRTASVIALTDGELHE	155		
DB	98	GSGGDVIR-AIRELS--YKGNRTGNAILHVADHVPOLARPGVPKVCILITDQK-SQ	153		
QY	156	DLFFYSERENSRDLGAIYVCGVDFNETQLARIAD--SKDHVPVNDGPFALQGIH	213		
DB	154	DLV---DTAAQLKGGCVKLFVAGIKNADPEELKRVASQPTSDFFVFND-FSILRTLPP	209		
QY	214	SILKKSCL-----EILAAEPST	230		
DB	210	LVSRRVCTTAGGVVTRPPDDSTSAPRDLVLSEPS	245		

Search completed: August 9, 2002, 10:34:21  
 Job time: 260 sec

.....

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:32:13 ; Search time 119.13 Seconds  
(without alignments)  
310.481 Million cell updates/sec

Title: US-09-970-076-8  
Perfect score: 1728  
Sequence: 1 MATERRALGIGFQWLSLAT.....TTHCSLHKASGPTTAACME 333

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_032802:\*  
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1728	100.0	333	21 AAB01422	Human TANGO 197
2	1649	95.4	403	22 AAE01439	Human gene 4 encod
3	1644	95.1	403	22 AAE01469	Human gene 4 encod
4	1392	80.6	297	22 AAM38976	Human polypeptide
5	966	55.9	218	22 AAB92985	Human protein sequ
6	801.5	46.4	488	21 AAB18456	A human TANGO 216
7	800.5	46.3	488	21 AAB18447	Amino acid sequenc
8	798.5	46.2	488	21 AAB18455	A human TANGO 216
9	797.5	46.2	488	21 AAB18457	A human TANGO 216
10	785.5	45.5	587	22 AAU19662	Human novel extrac
11	772	44.7	487	21 AAB18458	A murine TANGO 216

12	771	44.6	487	21 AAB18448	Amino acid sequenc
13	770	44.6	487	21 AAB18460	A murine TANGO 216
14	768	44.4	487	21 AAB18459	A murine TANGO 216
15	669	38.7	381	21 AAB01428	Murine TANGO 197
16	205	11.9	93	22 AAM84307	Human immune/haema
17	166	9.6	1155	16 AAR78167	Mouse alpha-d subu
18	166	9.6	1155	18 AAW23060	Mouse beta 2 integ
19	166	9.6	1155	19 AAW72835	Mouse alpha-d #1.
20	166	9.6	1155	19 AAW65102	Mouse beta-integri
21	166	9.6	1155	19 AAW60002	Mouse alpha d poly
22	166	9.6	1155	20 AAW73346	Mouse alphad prote
23	166	9.6	1155	21 AAB07372	Mouse alpha-d subu
24	166	9.6	1161	16 AAR78168	Mouse beta 2 integ
25	166	9.6	1161	18 AAW23061	Mouse alpha-d #2.
26	166	9.6	1161	19 AAW72836	Mouse beta-integri
27	166	9.6	1161	19 AAW65103	Mouse alpha d poly
28	166	9.6	1161	19 AAW60003	Mouse alphad prote
29	166	9.6	1161	20 AAW73347	Mouse alpha-d subu
30	166	9.6	1161	21 AAB07373	Rat beta 2 integri
31	159.5	9.2	1151	18 AAW23059	Rat alpha-d #2. R
32	159.5	9.2	1151	19 AAW72834	Rat beta-integri
33	159.5	9.2	1151	19 AAW65101	Rat alpha d polype
34	159.5	9.2	1151	19 AAW60001	Rat alphad protein
35	159.5	9.2	1151	20 AAW73344	Rat alpha-d partia
36	159.5	9.2	1151	21 AAB07371	Rat alpha-d subuni
37	159.5	9.2	1161	16 AAR78169	Human beta-2 integ
38	159.5	9.2	1161	16 AAR78166	Rat beta 2 integri
39	159.5	9.2	1161	18 AAW23082	Human beta 2 integ
40	159.5	9.2	1161	18 AAW23064	Human beta 2 integ
41	159.5	9.2	1161	18 AAW23049	Human alpha-d deri
42	159.5	9.2	1161	19 AAW72837	Human alpha-d #1. R
43	159.5	9.2	1161	19 AAW72824	Human alpha-d. Ho
44	159.5	9.2	1161	19 AAW72825	Human Beta-integri
45	159.5	9.2	1161	19 AAW65106	

#### ALIGNMENTS

RESULT 1  
AAB01422  
ID AAB01422 standard; Protein; 333 AA.

XX AC AAB01422;

XX DT 20-OCT-2000 (first entry)

XX DE Human TANGO 197.

XX KW TANGO: 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;  
XX KW graft versus-host diseases; rheumatoid arthritis; psoriasis;  
XX KW inflammatory bowel disease; septic shock; ulcerative colitis;  
XX KW Crohn's disease; chronic myelogenous leukemia; cancer; liver  
XX KW disease; Hodgkin's disease; osteoarthritis; Lyme's disease;  
XX KW cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;  
XX KW systemic lupus erythematosus; transgenic animal; diagnosis;  
XX KW prognosis; prophylactic; therapeutic; human.

XX OS Homo sapiens.

XX PN WO200039284-A1.

XX PD 06-JUL-2000.

XX PF 23-DEC-1999; 99WO-US31025.

XX PR 30-DEC-1998; 98US-0223546.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Holtzman DA;

XX DR WPI; 2000-465743/40.

cardiovascular disorder; angiogenic disorder; kidney disorder;  
gastrointestinal disorder; pregnancy-related disorder;  
endocrine disorder; infection; wound healing; vulnery;  
cell culture; chemotaxis; food additive; gene therapy;  
binding partner identification; chromosome 19.

Homo sapiens.

Key 1..27 Location/Qualifiers  
Peptide /label= Signal\_peptide  
Protein 28..403  
/note= "Mature human secreted protein"

WO200134626-A1.

17-MAY-2001.

01-NOV-2000; 2000WO-US30045.

05-NOV-1999; 99US-0163581.

30-JUN-2000; 2000US-0215133.

(HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Komatsoulis GA, Moore PA, Birse CE, Ni J;

WPI; 2001-308778/32.

N-PSDB; AAD05303.

New nucleic acid molecules encoding 28 human secreted proteins for  
diagnosing, preventing, treating or ameliorating medical conditions and  
used as food additives or preservatives

Claim 11; Page 485-486; 562pp; English.

AAD05300-AAD05379 represent cDNAs corresponding to 28 human secreted  
protein genes, and AAE01436-AAE01513 represent the proteins they encode.  
AAE01514-AAE01544 represent human secreted protein fragments or variants.  
The genes and their secreted proteins are useful for preventing,  
treating or ameliorating medical conditions, e.g., by protein or gene  
therapy. Pathological conditions can be diagnosed by determining the presence of  
mutations in the new genes. Specific uses are described for each of the  
28 genes, based on the tissues in which they are most highly expressed,  
and include developing products for the diagnosis or treatment of  
proliferative disorders, cancer, tumours, foetal and developmental  
abnormalities, haematopoietic disorders, diseases of the immune system,  
AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
allergies, neurological disorders (e.g., Alzheimer's disease,  
Parkinson's disease), cognitive disorders, sepsis, diabetes, atherosclerosis,  
skin disorders (e.g., psoriasis), angioec disorders, kidney disorders,  
cardiovascular disorders, pregnancy-related disorders, endocrine  
gastrointestinal disorders, The proteins can also be used to aid wound  
healing and epithelial cell proliferation, to prevent skin aging due to  
sunburn, to maintain organs before transplantation, for supporting cell  
culture of primary tissues, to regenerate tissues, to identify their  
cognate ligands or binding partners, and in chemotaxis, and can be used  
as a food additive or preservative to modify storage properties.  
Antibodies specific for a protein of the invention can be used in  
in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
immunosorbent assay (ELISA). The present sequence represents a human  
secreted protein of the invention.

Sequence 403 AA;

Query Match 95.4%; Score 1649; DB 22; Length 403;

Best Local Similarity 100.0%; Pred. No. 2,1e-168;

Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

N-PSDB; AAA47455.  
Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,  
224 and 239 polypeptides useful for the treatment of asthma, rheumatoid  
arthritis, psoriasis and autoimmune diseases

Claim 8; Fig 4; 209pp; English.

Nucleic acids encoding TANGO polypeptides are useful as modulating  
agents for regulating cellular processes like asthma, graft  
versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory  
bowel disease, septic shock, ulcerative colitis, Crohn's disease,  
chronic myelogenous leukemia, cancer, liver disease, Hodgkin's  
disease, osteoarthritis, Lyme's disease, cachexia and autoimmune  
diseases e.g. myasthenia gravis, autoimmune diabetes and systemic  
lupus erythematosus. The nucleic acids are also useful for producing  
transgenic animals and the TANGO polypeptides themselves. Partial  
TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in  
forensic biology, for diagnostic assays, prognostic assays,  
pharmacogenomics and for monitoring clinical trials. TANGO  
polypeptides are suitable for both prophylactic and therapeutic  
methods for treating a subject at risk of a disorder or having a  
disorder associated with aberrant TANGO expression. A wide range  
of cellular disorders can be treated.

Sequence 333 AA;

Query Match 100.0%; Score 1728; DB 21; Length 333;  
Best Local Similarity 100.0%; Pred. No. 5e-177;  
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MATAERRALGIGFOWLSATLVLCAGGGREDGGPACYGFDLYFLDKSGSVLHWN 60

1 metaerralgigfowlsatlvlicagggredggpacygfgdlyflldksgsvlhwn 60

61 EYIYFVEQLAHFISPOLRMSFIVSTRGTTLMKLTEDREIQGLELOKVLPGDFTY 120

61 eiyfyveqlahfispqlrmsfivstrgttlmkltedreqgleelqkvlpggdtym 120

121 HEGFERASEQIYYENRQGYRTASVIALTDGELHEDLFYSEANRSDIGAIYVCV 180

121 hegferaseqiyyenrqqyrtasvialtdgelhedllyfseanrnsrdigaiyvcv 180

181 KDFNETQLARIADSKDHPVNDGFOALQIIHSILKKSCIEILAAEPSTICAGESF 240

181 kdfnetqlariadskdhvfpvndgfgalqgiihsilkkscieilaepsticagesf 240

241 VRNGFRHARNVDRLCSFKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQV 300

241 vrngfrharnvdrvlcsfkindsvtlnekpsvedtyllcpapilkevgmkaalqv 300

301 DGLSFTSSSVIIITTHCSLHKIASGPTTAACME 333

301 dglssftsssviitthcslnhkiassgpttaacme 333

RESULT 2

AAE01439

ID AAE01439 standard; Protein; 403 AA.

XX AAE01439;

17-JUL-2001 (first entry)

Human gene 4 encoded secreted protein HHLFR02, SEQ ID NO:94.

Human; secreted protein; proliferative disorder; cancer; tumour;  
foetal abnormality; developmental abnormality; haematopoietic disorder;  
immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
inflammation; allergy; neurological disorder; Alzheimer's disease;  
Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;

```
QY 1 MATAERRALGIGFQWLSLATLVLCAGOGGRRDGGPACYGFGDLYFILDKSGSVLHHWN 60
PR |
PR |
XX |
Db 1 mataerralgigfqlslatlvlicagggrrdggpacyggfdlyfldksgsvlhhwn 60
QY 61 EIIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQIRGLEELQKVLPGGDTYM 120
PR |
PR |
XX |
Db 61 eiiyfyveqlahkfispqlrmsfivstrgttlnkltedreqirggleelqkvlpggdtym 120
QY 121 HEGFERASEQIYYENRQGYRTASVITATLDGELHEDLFFYSERANRSLDGLAIYVCVGV 180
PR |
PR |
XX |
Db 121 hegferaseqiiyenrqqyrtasvialtdgelhedlffysereanrslrdglaiyvcgv 180
QY 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHSLIKKSCIEILAAEPSTICAGESFQVW 240
PR |
PR |
XX |
Db 181 kdfnetqlariadskdhvfpvndgfqalqgihslikkscieilaepsticagesfvqw 240
QY 241 VRGNGFRHARNVDRVLCSPKINDSVTLNEKPFESVEDTYLLCPAPILKEVGMKAALQVSMN 300
PR |
PR |
XX |
Db 241 vrgngfrharnvdrvlcsfkindsvtlnkpfsvdtyllcpapilkevgmkaalqvsmn 300
QY 301 DGLSPFISSSVIITTHCS 318
PR |
PR |
XX |
Db 301 dglspfiSSSVIITTHCS 318
RESULT 3
ID AAE01469 standard; Protein; 403 AA.
AC AAE01469;
XX
DT 17-JUL-2001 (first entry)
DE Human gene 4 encoded secreted protein HWLFR02, SEQ ID NO:125.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; kidney disorder;
KW cardiovascular disorder; angiogenic disorder; atherosclerosis;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vunerary;
KW cell culture; chemotaxis; food additive; gene therapy;
KW binding partner identification; chromosome 19.
XX
OS Homo sapiens.
XX
Key Location/Qualifiers
PH Peptide 1..27
FT /label= Signal_peptide
FT Protein 28..403
FT /note= "Mature human secreted protein"
FT Misc-difference 175
FT /label= Unknown
FT /note= "Encoded by GKT"
FT Misc-difference 320
FT /label= Unknown
FT /note= "Encoded by SGT"
FT Misc-difference 331
FT /label= Unknown
FT /note= "Encoded by KTC"
FT Misc-difference 368
FT /label= Unknown
FT /note= "Encoded by WGC"
XX
WO200134626-A1.
XX
17-MAY-2001.
XX
01-NOV-2000; 2000WO-US30045.
XX
```

```
PR 05-NOV-1999; 99US-0163581.
PR 30-JUN-2000; 2000US-0215133.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Komatsoulis GA, Moore PA, Birse CE, NI J;
XX WPI: 2001-308778/32.
XX N-PSDB; AAD05334.
DR
DR
XX New nucleic acid molecules encoding 28 human secreted proteins for
XX diagnosing, preventing, treating or ameliorating medical conditions and
XX used as food additives or preservatives -
XX Claim 11; Page 505-506; 562pp; English.
XX
AA05300-AAD05379 represent cDNAs corresponding to 28 human secreted
CC protein genes, and AAE01436-AAE01513 represent the proteins they encode.
CC AAE01514-AAE01544 represent human secreted protein fragments or variants.
CC The genes and their secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 28 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein of the invention.
XX
SQ Sequence 403 AA;
Query Match 95.1%; Score 1644; DB 22; Length 403;
Best Local Similarity 99.7%; Pred No. 7, 3e-168;
Matches 317; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MATAERRALGIGFQWLSLATLVLCAGOGGRRDGGPACYGFGDLYFILDKSGSVLHHWN 60
Db |
Db 1 mataerralgigfqlslatlvlicagggrrdggpacyggfdlyfldksgsvlhhwn 60
QY 61 EIIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQIRGLEELQKVLPGGDTYM 120
Db |
Db 61 eiiyfyveqlahkfispqlrmsfivstrgttlnkltedreqirggleelqkvlpggdtym 120
QY 121 HEGFERASEQIYYENRQGYRTASVITATLDGELHEDLFFYSERANRSLDGLAIYVCVGV 180
Db |
Db 121 hegferaseqiiyenrqqyrtasvialtdgelhedlffysereanrslrdglaiyvcgv 180
QY 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHSLIKKSCIEILAAEPSTICAGESFQVW 240
Db |
Db 181 kdfnetqlariadskdhvfpvndgfqalqgihslikkscieilaepsticagesfvqw 240
QY 241 VRGNGFRHARNVDRVLCSPKINDSVTLNEKPFESVEDTYLLCPAPILKEVGMKAALQVSMN 300
Db |
Db 241 vrgngfrharnvdrvlcsfkindsvtlnkpfsvdtyllcpapilkevgmkaalqvsmn 300
```

Query Match 80.6%; Score 1392; DB 22; Length 297;  
 Best Local Similarity 99.3%; Pred. No. 5.9e-141;  
 Matches 267; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATAERRALGIGFOWLSLATLVLCAGGGRREDGGPACYGDFLYFILDKSGSVLHHWN 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 metaerralgigfowlslatlvlicagggrrredggpacygdflyfildksgsvllhwn 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 61 EIIYFEVQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLELQKVLPGGDTYM 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 61 eiyyfveqlahkfispolrmsfivfstrgttlmkltdedreqirggleelqkvlpggdtym 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 121 HEGFERASEQIYYENRQGYRTASVITIALTDGELHEDLFFYSERANRSDLGAIYVCVG 180  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 121 hegferaseqiyyenrqgyrtasvialtdgelhedlffysereanrdsrdigaivvcvg 180  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 181 KPFNETQIARIADSKDHPVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFOV 240  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 181 kpfnetqiaridaskdhvfvndgfoalqglihslkscieilaepsticagesfgv 240  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 241 VRGNGFRHARNVDRVLCSEKINDSVTLNE 269  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 241 vrgngfrharnvdrvclscfkindsvtlsk 269  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 5  
 AAB92985  
 ID AAB92985 standard; Protein; 218 AA.  
 XX  
 AC AAB92985;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:11706.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 11706; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end

QY 301 DGLSFSSVLIITTHCS 318  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 301 dglsfssvliitthcs 318  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 4  
 AAM38976  
 ID AAM38976 standard; Protein; 297 AA.  
 XX  
 AC AAM38976;  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 2121.  
 XX  
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 26-DEC-2000; 2000WO-US34263.  
 XX  
 PR 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI; 2001-442253/47.  
 DR N-PSDB; AAI58132.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 XX  
 PS Example 4; SEQ ID NO 2121; 10078pp; English.  
 XX  
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX  
 SQ Sequence 297 AA;



CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX Sequence 218 AA;

Query Match 55.9%; Score 966; DB 22; Length 218;  
 Best Local Similarity 98.9%; Pred. No. 2.5e-95;  
 Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 80 MSFVSTGTTLKMTEDREIQRLQLELQKVLPGDYMHEGFRASEQIYYENRQGY 139  
 Db 1 msfivstgrrtlmkitedreqirggleelqkvlpggdtymhegferaseqiyyenrgy 60  
 QY 140 RTASVLIATDGLHEDLEFYSEANRSDILGAIYVCVGRDNETQLARTADSKDHFV 199  
 Db 61 rtasvliatdglhedllyfseanrdsrdilgavycvgrdkfnetqlartadskdhvf 120  
 QY 200 PVNDGFOALQGIHTHSLKSKTEILAAEPSTICAGSFQVVRGNGFRHARNVDRLVCSF 259  
 Db 121 pvndgfoalqgihtslkskscieilaepsticagsfqvvrgngfrharnvdrlvcsf 180  
 QY 260 KINDSVTLNE 269  
 Db 181 kindsvtlslk 190

RESULT 6  
 AAB18456  
 ID AAB18456 standard; Protein; 488 AA.

XX AAB18456;

XX 15-JAN-2001 (first entry)

XX A human TANGO 216 polypeptide clone.

XX TANGO 266; TANGO 216; TANGO 261; TANGO 262; TANGO 267;  
 KW cellular proliferation; cellular differentiation; cellular adhesion;  
 KW von Willebrand factor-associated disorder; cell trafficking; cancer;  
 KW hematopoietic associated disease; atelectasis; pulmonary congestion;  
 KW oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;  
 KW intestinal disorder; spleen associated disease; renal disorder;  
 KW cardiovascular disorder; ischemic heart disease; hydrocephalus;  
 KW brain herniation; iatrogenic disease; inflammation; meningitis;  
 KW Alzheimer's Disease; cerebral toxoplasmosis; Parkinson's disease;  
 KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.

XX Homo sapiens.

OS WO200052022-A1.

PN 08-SEP-2000.

XX 01-MAR-2000; 2000WO-US05226.

XX 01-MAR-1999; 99US-0122458.

XX (MILL-) MILLENNIUM PHARM INC.

XX Barnes TM, Holtzman DA, Sharp JD, Fraser CC;

XX  
 DR  
 DR  
 XX

PT WPI: 2000-579269/54.  
 PT N-PSDB; AAA75158.  
 PT Novel human and murine secreted proteins designated TANGO 216, 261,  
 PT 262, 266 and 267 useful as modulating agents of cellular processes,  
 PT e.g. for treating cancer -

XX Disclosure: Page -; 175pp; English.

XX AAB18455-57 represent human TANGO 216 proteins. The specification also  
 CC describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO  
 CC polypeptides can be used to modulate cellular proliferation, modulate  
 CC cellular differentiation and/or modulate cellular adhesion. The  
 CC proteins can be used to treat any von Willebrand factor-associated  
 CC disorder, regulate extracellular matrix structuring, cellular adhesion,  
 CC and cell trafficking and/or migration, modulate cellular interactions,  
 CC modulate cell adhesion in proliferative disorders, such as cancer,  
 CC modulate the proliferation, differentiation, and/or function of cells  
 CC that appear in the bone marrow, and leukocytes, treat bone marrow, blood  
 CC and hematopoietic associated diseases and disorders, atelectasis,  
 CC pulmonary congestion or oedema, emphysema, chronic bronchitis, bronchial  
 CC asthma and bronchiectasis, intestinal disorders, spleen associated  
 CC diseases, modulate renal disorders, treat cardiovascular disorders such  
 CC as ischemic heart disease, modulate the proliferation, differentiation,  
 CC and/or function of bone and cartilage cells and to treat bone and/or  
 CC cartilage associated diseases or disorder. They may also be used to  
 CC treat disorders associated with the ovaries, and cerebral oedema,  
 CC hydrocephalus, brain herniations, iatrogenic disease, inflammations,  
 CC bacterial and viral meningitis, Alzheimer's disease, cerebral  
 CC toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers,  
 CC hydrocephalus and encephalitis, and treat hepatic disorders.  
 CC note: the present sequence does not appear in the specification; it was  
 CC created using information provided.

XX Sequence 488 AA;

Query Match 46.4%; Score 801.5; DB 21; Length 488;  
 Best Local Similarity 51.3%; Pred. No. 4.2e-77;  
 Matches 162; Conservative 56; Mismatches 95; Indels 3; Gaps 2;

QY 4 AERRALGIGFQWLSLATLVLCAGQG-RREDGGPACYGDFLYFLDKSGSVLHHWNEI 62  
 Db 3 aerspaspqswlfpqwlwlvlsqpgllraqdqpcrrafdlvldksgsvannwnei 62

QY 63 YFVEOLAHKFTSPQLRMSFIVFSTRGTTLMKLTEDREIQRLQLELQKVLPGDYMHE 122

Db 63 ynfvqqlaerfvspemrlsfivssqatillptgdrqkiskgledlkrvspvgetyie 122

QY 123 GPERASEQIYYENRQGYRTASVLIATDGLHEDLEFYSEANRSDILGAIYVCVGRD 182

Db 123 giklaneql--qkaglkssliialtdqkldglvpsyaekaklsrsagsvcyvgvld 180

QY 183 FNETQLARIADSKDHFVPPVNDGFOALQGIHTHSLKSKTEILAAEPSTICAGSFQVVR 242

Db 181 feaqlerladskeqvfpvkggqalkginsilaqscleilelqpsvscvgeefqlvs 240

QY 243 GNGFRHARNVDRLVCSFKINDSVTLNEKPFSEVEDTLLCPAPILKEVGMKAALQVSMNDG 302

Db 241 grgfmlgsrnsgsvlctytnetytsvkpsvqlnsmldcpapilnkgateldvsvsfngg 300

QY 303 LSPFISSSVIITTHCS 318

Db 301 ksvigsllvatecs 316

RESULT 7

AAB18447

ID AAB18447 standard; Protein; 488 AA.

XX AAB18447;

XX





Fri Aug 9 10:57:06 2002

```

QY 63 YFVQLAHKTFSPQIRMSFVIFSTRGTTMLKLTEDREOIROGLELQKVLPGDGYMHE 122
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 63 ynfvgqlaerfvspeirfsvfssqatiipldtdrkiskgledlkrvsvpgetyih 122
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 123 GPERASEQIYYENRQYRTASVITIALTDGELHEDLFFYSERANKSRDLGAIYVCVGVKD 182
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 123 giklaneql--qkagglkssialtdgldgklypsyaekakrsrsgasvycgvld 180
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 183 FNETQIARADSKDHVPFNDGQALQGIHSLTLAKKSCIEILAAPSTICAGESFQVWVR 242
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 181 feqaqleriadskeqfvpkggfgalknslsllaqscceillelqpsvvcveeqfivls 240
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 243 GNGFRHARNVDRLVCSFKINDSVTLNEKFPFVEDTYLLCPAPILKEVGMKAALQVSMNDG 302
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 241 grgfmjgsrngsvlctytnvetttsvkpsvqlnsmllcpapilnkagetldvsvsfngg 300
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY, 303 LSFSSSVIITTHCS 318
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 301 ksvlsgslivtatecs 316
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 10
AAU19662
ID AAU19662 standard; Protein; 587 AA.
AC AAU19662;
XX
DT 06-DEC-2001 (first entry)
DE Human novel extracellular matrix protein, Seq ID No 312.
KW Human; secreted extracellular matrix protein; immunomodulatory;
KW Anti-HIV; antianemic; antirheumatic; antisclerotic; cardiac; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytostatic;
KW antiaizheimers; immune/autoimmune disease; HIV infection; anaemia;
KW human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;
KW cancers; hyperproliferative disorder; breast neoplasia; melanoma;
KW Sezary syndrome; Gaucher's disease; neurological diseases;
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
KW cardiac arrest; tachycardia; angina; infection; corneal infections;
KW wound healing; immunogen; gene therapy; antisense; food additive.
XX Homo sapiens.
OS
XX WO20015368-A1.
PN
XX
PD
XX 02-AUG-2001.
PF 17-JAN-2001; 2001WO-US01348.
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205123.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 01-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.

```



PT Novel human and murine secreted proteins designated TANGO 216, 261, 262, 266 and 267 useful as modulating agents of cellular processes, e.g. for treating cancer -

PS Disclosure: Page -: 175pp; English.

XX AAB18458-60 represent murine TANGO 216 proteins. The specification also describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO polypeptides can be used to modulate cellular proliferation, modulate cellular differentiation and/or modulate cellular adhesion. The proteins can be used to treat any von Willebrand factor-associated disorder, regulate extracellular matrix structuring, cellular adhesion, and cell trafficking and/or migration, modulate cellular interactions, modulate cell adhesion in proliferative disorders, such as cancer, modulate the proliferation, differentiation, and/or function of cells that appear in the bone marrow, and leukocytes, treat bone marrow, blood and hematopoietic associated diseases and disorders, atelectasis, pulmonary congestion or oedema, emphysema, chronic bronchitis, bronchial asthma and bronchiectasis, intestinal disorders, treat cardiovascular disorders such as ischemic heart disease, modulate the proliferation, differentiation, and/or function of bone and cartilage cells and to treat bone and/or cartilage associated diseases or disorder. They may also be used to treat disorders associated with the ovaries, and cerebral oedema, hydrocephalus, brain herniations, iatrogenic disease, cerebral toxoplasmosis, Parkinson's Disease, brain cancers, bacterial and viral meningitis, Alzheimer's Disease, cerebral toxoplasmosis, Parkinson's disease, multiple sclerosis, and treat hepatic disorders. CC note: the present sequence does not appear in the specification; it was created using information provided.

XX Sequence 487 AA;

Query Match 44.7%; Score 772; DB 21; Length 487;  
Best Local Similarity 49.5%; Pred. NO. 6.3e-74;  
Matches 158; Conservative 53; Mismatches 104; Indels 4; Gaps 3;

QY 1 MATAERRAIGIGFQWLSLATLVLCAGGG-RREDGGPACYGGEFDLYFLDKSGSVLHHW 59  
Db 1 mvagrsrarspg-swlfpglwllavggpslllgaqdpckkafdyfldksgsvannw 59  
QY 60 NEIYVEQIAHFKFISPOLRMSFIVFSTRGTTLMKLTEDREQIRGLEELQKVLPGGDTY 119  
Db 60 ielnfvhlterfvspemlrsfivssqatillptgrykigkqledlkavkpvgety 119  
QY 120 MHGFFERASEQIYYENRQGYRTASVIALTDGELHEDLFFYSERANRSDLGAIYVCVG 179  
Db 120 iheglklaneqi--qngaglikassilialtdgklglvpsyaeneakrsigasyvcvg 177  
QY 180 VKDFNETQIARIADSKDHFVVDNGFQALQGIHISILKKSCBILAAEPSTICAGESFOV 239  
Db 178 vldfeaqqlerladskdqfvpkggfgalqinsilaqscleilelsspsvcvgekfv 237  
QY 240 VVRNGCFRHRARNVDRLVCSFKINDSVTLNKPFSVEDTYLLCPAPILKEVGMKAALQVSM 299  
Db 238 vltgravtsishdgsvlctftanstytksekpsvsgpslcpapvlnkdgetlevisy 297  
QY 300 NDGLSFSSSVITTHCS 318  
Db 298 ndgksavrsititact 316

RESULT 12  
AAB18448  
ID AAB18448 standard; Protein; 487 AA.  
AC AAB18448;  
XX  
DT 15-JAN-2001 (first entry)  
XX  
DE Amino acid sequence of a murine TANGO 216 polypeptide.

KW TANGO 266; TANGO 216;; TANGO 261; TANGO 262; TANGO 267;  
KW cellular proliferation; cellular differentiation; cellular adhesion;  
KW von Willebrand factor-associated disorder; cell trafficking; cancer;  
KW hematopoietic associated disease; atelectasis; pulmonary congestion;  
KW oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;  
KW intestinal disorder; spleen associated disease; renal disorder;  
KW cardiovascular disorder; ischemic heart disease; hydrocephalus;  
KW brain herniation; iatrogenic disease; inflammation; meningitis;  
KW Alzheimer's Disease; cerebral toxoplasmosis; Parkinson's disease;  
KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.  
XX  
XX Mus sp.  
OS  
XX  
FH Location/Qualifiers  
FT Key 1..33  
FT Peptide /note= "signal peptide"  
FT Domain 34..79 /note= "extracellular domain"  
FT Protein 34..487 /note= "mature protein"  
FT Domain 80..97 /note= "transmembrane domain"  
FT Domain 98..317 /note= "cytoplasmic domain"  
FT Domain 318..341 /note= "transmembrane domain"  
FT Domain 342..487 /note= "extracellular domain"  
XX  
PN W0200052022-AL.  
XX  
XX 08-SEP-2000.  
XX  
XX 01-MAR-2000; 2000WO-US05226.  
XX  
XX 01-MAR-1999; 99US-0122458.  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
XX  
XX Barnes TM, Holtzman DA, Sharp JD, Fraser CC;  
XX  
XX WPI: 2000-579269/54.  
XX  
XX N-PSDB; AAA75150.  
XX  
XX Novel human and murine secreted proteins designated TANGO 216, 261,  
XX 262, 266 and 267 useful as modulating agents of cellular processes,  
XX e.g. for treating cancer -  
XX  
XX Claim 8; Fig 2A-C; 175pp; English.  
XX  
XX The present sequence represents a murine TANGO 216 polypeptide. The  
XX specification also describes TANGO 266, TANGO 261, TANGO 262, and  
XX TANGO 267. The TANGO polypeptides can be used to modulate cellular  
XX proliferation, modulate cellular differentiation and/or modulate  
XX cellular adhesion. The proteins can be used to treat any von Willebrand  
XX factor-associated disorder, regulate extracellular matrix structuring,  
XX cellular adhesion, and cell trafficking and/or migration, modulate  
XX cellular interactions, modulate cell adhesion in proliferative  
XX disorders, such as cancer, modulate the proliferation, differentiation,  
XX and/or function of cells that appear in the bone marrow, and leukocytes,  
XX treat bone marrow, blood and hematopoietic associated diseases and  
XX disorders, atelectasis, pulmonary congestion or oedema, emphysema,  
XX chronic bronchitis, bronchial asthma and bronchiectasis, intestinal  
XX disorders, spleen associated diseases, modulate renal disorders, treat  
XX cardiovascular disorders such as ischemic heart disease, modulate the  
XX proliferation, differentiation, and/or function of bone and cartilage  
XX cells and to treat bone and/or cartilage associated diseases or  
XX disorder. They may also be used to treat disorders associated with the  
XX ovaries, cerebral oedema, hydrocephalus, brain herniations, iatrogenic  
XX disease, inflammation, bacterial and viral meningitis, Alzheimer's  
XX disease, cerebral toxoplasmosis, Parkinson's disease, multiple sclerosis,  
XX brain cancers, hydrocephalus and encephalitis, and treat hepatic  
XX disorders.

```

XX SQ Sequence 487 AA;
Query Match 44.6%; Score 771; DB 21; Length 487;
Best Local Similarity 49.5%; Pred. No. 8.1e-74;
Matches 158; Conservative 53; Mismatches 104; Indels 4; Gaps 3;

QY 1 MATERRALGIGFOWLSLATLVLCACGGG--RREDGGPACVGGDFLYFLDKSGSVLHHW 59
Db 1 mvagrsrarspg-swlfpgllwllavggpgsllqaeqpsckkaaflyfvlksgsvannw 59
QY 60 NEIYFVEQLAHKFTISPOLRMSFVFSRGTTLMLKLTEDREQIRQGLEELQKVLPGGDTY 119
Db 60 leiynfvhqlterfvspeimrslfvsqsatliilptgdrykigkgleedlkavkpvgety 119
QY 120 MHGCFERASEQIYENROGYTASVITALTDCGLHEDLFFYSERANRDLGATVYCVG 179
Db 120 ihlegiklaneqi--qnaaglkassilialtdgklidglvpsyaeneakksrslgasvycv 179
QY 180 VKDFNETQLARIADSKDHVPVNDGFQALOGIIHSILKKSCTEILAAPSTICAGESFQV 239
Db 178 vldfeqaqleriadskdqvfpgkqfalkglnslagsctellelspssvcvgektqv 237
QY 240 VVRNGFRHARNVDRVLCFSKINDSVTLNRPFSVEDTYLLCPAPILKEVGMKAALQVSM 299
Db 238 vltgravtsishdgsvlctftanstytksekpvsilqpsilcpavlnkdgetlevsisy 297
QY 300 NDGLSFSSSVIITTHCS 318
Db 298 ndgksavrsrltitatect 316

RESULT 13
AAB18460
ID AAB18460 standard; Protein: 487 AA.
XX AC AAB18460;
XX DT 15-JAN-2001 (first entry)
XX DE A murine TANGO 216 polypeptide clone.
XX KW TANGO 266; TANGO 216;; TANGO 261; TANGO 262; TANGO 267;
KW cellular proliferation; cellular differentiation; cellular adhesion;
KW von Willebrand factor-associated disorder; cell trafficking; cancer;
KW hematopoietic associated disease; atelectasis; pulmonary congestion;
KW oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;
KW intestinal disorder; spleen associated disease; renal disorder;
KW cardiovascular disorder; ischemic heart disease; hydrocephalus;
KW brain herniation; iatrogenic disease; inflammation; meningitis;
KW Alzheimer's disease; cerebral toxoplasmosis; Parkinson's disease;
KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.
XX OS Mus sp.
XX PN WO200052022-A1.
XX PD 08-SEP-2000.
XX PF 01-MAR-2000; 2000WO-US05226.
XX PR 01-MAR-1999; 99US-0122458.
XX PA (MILL-) MILLENNIUM PHARM INC.
XX PI Barnes TM, Holtzman DA, Sharp JD, Fraser CC;
XX DR WPI; 2000-579269/54.
XX DR N-PSDB; AAA75162.
XX PT Novel human and murine secreted proteins designated TANGO 216, 261,
PT 262, 266 and 267 useful as modulating agents of cellular processes,

```

```

PT e.g. for treating cancer -
XX Disclosure; Page -: 175pp; English.
XX AAB18458-60 represent murine TANGO 216 proteins. The specification also
CC describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO
CC polypeptides can be used to modulate cellular proliferation, modulate
CC cellular differentiation and/or modulate cellular adhesion. The
CC proteins can be used to treat any von Willebrand factor-associated
CC disorder, regulate extracellular matrix structuring, cellular adhesion,
CC and cell trafficking and/or migration, modulate cellular interactions,
CC modulate cell adhesion in proliferative disorders, such as cancer,
CC modulate the proliferation, differentiation, and/or function of cells
CC that appear in the bone marrow, and leukocytes, treat bone marrow, blood
CC and hematopoietic associated diseases and disorders, atelectasis,
CC pulmonary congestion or oedema, emphysema, chronic bronchitis, bronchial
CC asthma and bronchiectasis, intestinal disorders, spleen associated
CC diseases, modulate renal disorders, treat cardiovascular disorders such
CC as ischemic heart disease, modulate the proliferation, differentiation,
CC and/or function of bone and cartilage cells and to treat bone and/or
CC cartilage associated diseases or disorder. They may also be used to
CC treat disorders associated with the ovaries, and cerebral oedema,
CC hydrocephalus, brain herniations, iatrogenic disease, inflammations,
CC bacterial and viral meningitis, Alzheimer's disease, cerebral
CC toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers,
CC hydrocephalus and encephalitis, and treat hepatic disorders.
CC note: the present sequence does not appear in the specification; it was
CC created using information provided.
XX SQ Sequence 487 AA;

Query Match 44.6%; Score 770; DB 21; Length 487;
Best Local Similarity 49.5%; Pred. No. 1e-73;
Matches 158; Conservative 53; Mismatches 104; Indels 4; Gaps 3;

QY 1 MATERRALGIGFOWLSLATLVLCACGGG--RREDGGPACVGGDFLYFLDKSGSVLHHW 59
Db 1 mvagrsrarspg-swlfpgllwllavggpgsllqaeqpsckkaaflyfvlksgsvannw 59
QY 60 NEIYFVEQLAHKFTISPOLRMSFVFSRGTTLMLKLTEDREQIRQGLEELQKVLPGGDTY 119
Db 60 leiynfvhqlterfvspeimrslfvsqsatliilptgdrykigkgleedlkavkpvgety 119
QY 120 MHGCFERASEQIYENROGYTASVITALTDCGLHEDLFFYSERANRDLGATVYCVG 179
Db 120 ihlegiklaneqi--qnaaglkassilialtdgklidglvpsyaeneakksrslgasvycv 177
QY 180 VKDFNETQLARIADSKDHVPVNDGFQALOGIIHSILKKSCTEILAAPSTICAGESFQV 239
Db 178 vldfeqaqleriadskdqvfpgkqfalkglnslagsctellelspssvcvgektqv 237
QY 240 VVRNGFRHARNVDRVLCFSKINDSVTLNRPFSVEDTYLLCPAPILKEVGMKAALQVSM 299
Db 238 vltgravtsishdgsvlctftanstytksekpvsilqpsilcpavlnkdgetlevsisy 297
QY 300 NDGLSFSSSVIITTHCS 318
Db 298 ndgksavrsrltitatect 316

RESULT 14
AAB18459
ID AAB18459 standard; Protein: 487 AA.
XX AC AAB18459;
XX DT 15-JAN-2001 (first entry)
XX DE A murine TANGO 216 polypeptide clone.
XX KW TANGO 266; TANGO 216;; TANGO 261; TANGO 262; TANGO 267;
KW cellular proliferation; cellular differentiation; cellular adhesion;

```

von Willebrand factor-associated disorder; cell trafficking; cancer; hematopoietic associated disease; atelectasis; pulmonary congestion; oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis; intestinal disorder; spleen associated disease; renal disorder; cardiovascular disorder; ischemic heart disease; hydrocephalus; brain herniation; iatrogenic disease; inflammation; meningitis; Alzheimer's Disease; cerebral toxoplasmosis; Parkinson's disease; multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.

Mus sp.

WO200052022-A1.

08-SEP-2000.

01-MAR-2000; 2000WO-US05226.

01-MAR-1999; 99US-0122458.

(MILL-) MILLENNIUM PHARM INC.

Barnes TM, Holtzman DA, Sharp JD, Fraser CC;

WPI; 2000-579269/54.

N-PSDB; AAA75161.

Novel human and murine secreted proteins designated TANGO 216, 261, 262, 266 and 267 useful as modulating agents of cellular processes, e.g. for treating cancer -

Disclosure; Page -: 175pp; English.

AAB18458-60 represent murine TANGO 216 proteins. The specification also describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO polypeptides can be used to modulate cellular proliferation, modulate cellular differentiation and/or modulate cellular adhesion. The proteins can be used to treat any von Willebrand factor-associated disorder, regulate extracellular matrix structuring, cellular adhesion, and cell trafficking and/or migration, modulate cellular interactions, modulate cell adhesion in proliferative disorders, such as cancer, modulate the proliferation, differentiation, and/or function of cells that appear in the bone marrow, and leukocytes, treat bone marrow, blood and hematopoietic associated diseases and disorders, atelectasis, asthma and bronchiectasis, intestinal disorders, spleen associated diseases, modulate renal disorders, treat cardiovascular disorders such as ischemic heart disease, modulate the proliferation, differentiation, and/or function of bone and cartilage cells and to treat bone and/or cartilage associated diseases or disorder. They may also be used to treat disorders associated with the ovaries, and cerebral oedema, hydrocephalus, brain herniations, iatrogenic disease, inflammation, bacterial and viral meningitis, Alzheimer's Disease, cerebral toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers, hydrocephalus and encephalitis, and treat hepatic disorders.

note: the present sequence does not appear in the specification; it was created using information provided.

Sequence 487 AA;

Query Match 44.4%; Score 768; DB 21; Length 487;

Best Local Similarity 49.2%; Pred. No. 1.7e-73;

Matches 157; Conservative 54; Mismatches 104; Indels 4; Gaps 3;

QY 1 MATARRALGIFGOWLSLALVILICAGQG--RREDGGPACYGFDLYFLIDKSGSVLHHW 59

Db 1 mvaqrarspg-swlfpgllavpgpgslllqeqpsckkafdyfvdksqsvannw 59

QY 60 NIIYFVEQLAKHTISPOLRMSFVFTSTLMKLTEDREQVROGLEELQKVLPGGDTY 119

Db 60 idyfnvqhlerfvspemlrsfivfsqatillptgdrykgkgledlkavkprgety 119

QY 120 MHEGFERASEQIYYENROGYRTASVITALTDELHEDLFFYSERANRSDLGAIIVCVG 179

Db 120 iheglklaneqi--qnaaglkassiiatdgdldgvlpsyaeneakksrsgasvycvg 177

QY 180 VKFNETQLARIADSKDHPVNDGFQALOGIIHSLKKSCIEILAAEPSTICAGESFOV 239

Db 178 vldfeqgleriadskdqvpvkggfgalkgiinsilaqscleilspssvcvgekfv 237

QY 240 VVRNGGFRHARNVDRVLCSPKINDSVTLNKPSPVEDTYLLCPAPILKEVGMKAALQVSM 299

Db 238 vltgravtsishdgsvlctftanstytksekpslqpsllcpapvinkdgetlevisy 297

QY 300 NDGLSFITSSSVIITTHCS 318

Db 298 ndgksavsrsltitatect 316

RESULT 15

AAB01428

ID AAB01428 standard; Protein; 381 AA.

XX

AC AAB01428;

DT 20-OCT-2000 (first entry)

XX

DE Murine TANGO 197.

XX

XX TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma; graft versus-host diseases; rheumatoid arthritis; psoriasis; inflammatory bowel disease; septic shock; ulcerative colitis; Crohn's disease; chronic myelogenous leukemia; cancer; liver disease; Hodgkin's disease; osteoarthritis; Lyme's disease; cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes; systemic lupus erythematosus; transgenic animal; diagnosis; prognosis; prophylactic; therapeutic; mouse.

XX Mus musculus.

OS WO200039284-A1.

XX

PN 06-JUL-2000.

PD

XX 23-DEC-1999; 99WO-US31025.

XX

PR 30-DEC-1998; 98US-0223546.

XX (MILL-) MILLENNIUM PHARM INC.

XX Holtzman DA;

XX

XX WPI; 2000-465743/40.

XX N-PSDB; AAA47479.

XX

XX Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213, 224 and 239 polypeptides useful for the treatment of asthma, rheumatoid arthritis, psoriasis and autoimmune diseases

XX

PS: Claim 8; Fig 27; 209pp; English.

XX

XX Nucleic acids encoding TANGO polypeptides are useful as modulating agents for regulating cellular processes like asthma, graft versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory bowel disease, septic shock, ulcerative colitis, Crohn's disease, chronic myelogenous leukemia, cancer, liver disease, Hodgkin's disease, osteoarthritis, Lyme's disease, cachexia and autoimmune diseases e.g. myasthenia gravis, autoimmune diabetes and systemic lupus erythematosus. The nucleic acids are also useful for producing transgenic animals and the TANGO polypeptides themselves. Partial TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in forensic biology, for diagnostic assays, prognostic assays, pharmacogenomics and for monitoring clinical trials. TANGO polypeptides are suitable for both prophylactic and therapeutic methods for treating a subject at risk of a disorder or having a disorder associated with aberrant TANGO expression. A wide range



CC of cellular disorders can be treated.

XX  
SQ Sequence 381 AA;

Query Match 38.7%; Score 669; DB 21; Length 381;  
Best Local Similarity 99.2%; Pred. No. 5.2e-63;  
Matches 131; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	187	QLARIADSKDHVFPVNDGFQALQGIHHSILKKSCIEILAAEPSTICAGESFQVVVRGNGF	246
Db	4	qlariadskdhvfpvndgfqalgglihsilkkscieilaaepsticagesfqvvvrngnf	63
Qy	247	RHARNVDRVLCSPKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGKKAALQVSMNDGLSFI	306
Db	64	rharndrvlcsfkindsvtlinekpfavedtyllcpapilkevgmkaalqvsmndglsfi	123
Qy	307	SSSVIITTHCS	318
Db	124	sssviittthcs	135

Search completed: August 9, 2002, 10:32:14  
Job time: 133 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:35:55 ; Search time 98.32 seconds  
(without alignments)  
1005.580 Million cell updates/sec

Title: US-09-970-076-6  
Perfect score: 3025  
Sequence: 1 MATERRALGIGFQWLSLAT.....QAPPNRPSPRPSPV 564

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 528882 seqs, 175299045 residues

Total number of hits satisfying chosen parameters: 528882

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_New:\*

- 1: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3025	100.0	564	1 PCT-US02-08253-187	Sequence 187, App
2	3025	100.0	564	1 PCT-US02-08253-232	Sequence 232, App
3	3025	100.0	564	7 US-60-389-987-1823	Sequence 1823, Ap
4	2901	95.9	562	1 PCT-US02-08253-194	Sequence 194, App
5	2901	95.9	562	1 PCT-US02-08253-301	Sequence 301, App
6	2017	66.7	381	6 US-10-038-307-4	Sequence 4, Appli
7	1661	54.9	551	6 US-10-038-307-18	Sequence 18, Appl
8	1649	54.5	333	6 US-10-038-307-2	Sequence 2, Appli
9	1649	54.5	345	6 US-10-038-307-24	Sequence 24, Appl
10	1649	54.5	564	6 US-10-038-307-20	Sequence 20, Appl
11	1636	54.1	328	6 US-10-038-307-26	Sequence 26, Appl
12	1634.5	54.0	342	6 US-10-038-307-22	Sequence 22, Appl
13	1629	53.9	543	6 US-10-038-307-14	Sequence 14, Appl
14	1629	53.9	543	6 US-10-038-307-16	Sequence 16, Appl
15	1619	53.5	543	6 US-10-038-307-10	Sequence 10, Appl
16	1488	49.2	534	6 US-10-038-307-12	Sequence 12, Appl
17	1416.5	46.8	488	6 US-10-104-047-2639	Sequence 2639, Ap
18	1411.5	46.7	488	6 US-10-038-307-6	Sequence 6, Appli
19	1411.5	46.7	488	6 US-60-373-595-10	Sequence 10, Appl
20	1396.5	46.2	487	6 US-10-038-307-8	Sequence 8, Appli
21	1206.5	39.9	587	6 US-10-125-540-312	Sequence 312, App
22	966	31.9	218	5 US-09-629-469A-11706	Sequence 11706, A
23	903	29.9	538	6 US-10-047-342-99	Sequence 99, Appl
24	211	7.0	1589	5 US-09-935-625-9406	Sequence 9406, Ap
25	211	7.0	1589	5 US-09-935-625-26246	Sequence 26246, A
26	211	7.0	1615	5 US-09-935-625-9405	Sequence 9405, Ap

27	211	7.0	1615	5 US-09-935-625-26245	Sequence 26245, A
28	211	7.0	1649	5 US-09-935-625-9404	Sequence 9404, Ap
29	211	7.0	1649	5 US-09-935-625-26244	Sequence 26244, A
30	199.5	6.6	669	5 US-09-935-625-9383	Sequence 9383, Ap
31	199.5	6.6	669	5 US-09-935-625-17852	Sequence 17852, A
32	199.5	6.6	681	5 US-09-935-625-9382	Sequence 9382, Ap
33	199.5	6.6	681	5 US-09-935-625-17851	Sequence 17851, A
34	199.5	6.6	715	5 US-09-935-625-9381	Sequence 9381, Ap
35	199.5	6.6	715	5 US-09-935-625-17850	Sequence 17850, A
36	194	6.4	592	5 US-09-935-625-900	Sequence 900, App
37	194	6.4	592	5 US-09-935-625-5343	Sequence 5343, Ap
38	194	6.4	592	5 US-09-935-625-8099	Sequence 8099, Ap
39	194	6.4	592	5 US-09-935-625-25320	Sequence 25320, A
40	194	6.4	597	5 US-09-935-625-899	Sequence 899, App
41	194	6.4	597	5 US-09-935-625-5342	Sequence 5342, Ap
42	194	6.4	597	5 US-09-935-625-8098	Sequence 8098, Ap
43	194	6.4	597	5 US-09-935-625-25319	Sequence 25319, A
44	194	6.4	727	5 US-09-935-625-898	Sequence 898, App
45	194	6.4	727	5 US-09-935-625-5341	Sequence 5341, Ap

## ALIGNMENTS

RESULT 1  
PCT-US02-08253-187  
; Sequence 187, Application PC/TUS0208253  
; GENERAL INFORMATION:  
; APPLICANT: Carson-Walter, Eleanor  
; APPLICANT: St. Croix, Brad  
; APPLICANT: Vogelstein, Bert  
; APPLICANT: Kinzler, Kenneth  
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
; FILE REFERENCE: 1107.00179  
; CURRENT APPLICATION NUMBER: PCT/US02/08253  
; CURRENT FILING DATE: 2002-04-10  
; PRIOR APPLICATION NUMBER: 60/282,850  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/308,829  
; PRIOR FILING DATE: 2001-08-01  
; NUMBER OF SEQ ID NOS: 359  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 187  
; LENGTH: 564  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US02-08253-187

Query Match		100.0%	Score 3025;	DB 1:	Length 564;
Best Local Similarity		100.0%	Pred. No. 3e-176;		
Matches 564;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MATERRALGIGFQWLSLATLVLCAGOGGREDGGPACYGDFLYFLDKSGVSLHHWN	60		
Db	1	MATERRALGIGFQWLSLATLVLCAGOGGREDGGPACYGDFLYFLDKSGVSLHHWN	60		
Qy	61	EIYFVEQLAHKFTSPQLRMSFIVFSTRGTTLMKLTEDREQIRQGLEELQKVLPGDITYM	120		
Db	61	EIYFVEQLAHKFTSPQLRMSFIVFSTRGTTLMKLTEDREQIRQGLEELQKVLPGDITYM	120		
Qy	121	HEGFERASEQIYYENRQGYRTASVITATDGLHEDLFFYSERANRSLDCAIYVCVGV	180		
Db	121	HEGFERASEQIYYENRQGYRTASVITATDGLHEDLFFYSERANRSLDCAIYVCVGV	180		
Qy	181	KDFNETQLARIADSKDHVFPVNDGFGALQGIHSLKKSCEILAAEPSTICAGESFQVW	240		
Db	181	KDFNETQLARIADSKDHVFPVNDGFGALQGIHSLKKSCEILAAEPSTICAGESFQVW	240		
Qy	241	VRNGFRHARNVDRVLCSEFKINDSVTLNEKPSVEDTLLCPAPILKEVGKAAALQVSMN	300		
Db	241	VRNGFRHARNVDRVLCSEFKINDSVTLNEKPSVEDTLLCPAPILKEVGKAAALQVSMN	300		

Fri Aug 9 10:57:05 2002

QY 301 DGLSFSSSVIIITTHCSGDSILAIALILFLLLALLLWFWPLCCTVIKEVPPPPAE 360  
Db 301 DGLSFSSSVIIITTHCSGDSILAIALILFLLLALLLWFWPLCCTVIKEVPPPPAE 360  
QY 361 ESEEDDDGLPKKWPPTVDASYGGGKIRMEVRWGEKSTEGAKLEKAKNARVKM 420  
Db 361 ESEEDDDGLPKKWPPTVDASYGGGKIRMEVRWGEKSTEGAKLEKAKNARVKM 420  
QY 421 PEQYEFPEPRNLNNMRSSPRKWTSPKGLDALWLLRKGYDRVSMRPPQDGTGR 480  
Db 421 PEQYEFPEPRNLNNMRSSPRKWTSPKGLDALWLLRKGYDRVSMRPPQDGTGR 480  
QY 481 CINFTRVKNQPAKYPLNNAYHTSSPPAPITYTPPPAPHCPPPPPSAPTPIPPSPSTL 540  
Db 481 CINFTRVKNQPAKYPLNNAYHTSSPPAPITYTPPPAPHCPPPPPSAPTPIPPSPSTL 540  
QY 541 PPPQAPPPNRAPPSPRPPPSV 564  
Db 541 PPPQAPPPNRAPPSPRPPPSV 564  
RESULT 2  
PCT-US02-08253-232  
; Sequence 232, Application PC/TUS0208253  
; GENERAL INFORMATION:  
; APPLICANT: Carson-Walter, Eleanor  
; APPLICANT: St. Croix, Brad  
; APPLICANT: Vogelstein, Bert  
; APPLICANT: Kinzler, Kenneth  
; TITLE OF INVENTION: ENDONUCLEIAL CELL EXPRESSION PATTERNS  
; FILE REFERENCE: 1107.00179  
; CURRENT APPLICATION NUMBER: PCT/US02/08253  
; CURRENT FILING DATE: 2002-04-10  
; PRIOR APPLICATION NUMBER: 60/282,850  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/308,829  
; PRIOR FILING DATE: 2001-08-01  
; NUMBER OF SEQ ID NOS: 359  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 232  
; LENGTH: 564  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US02-08253-232

Query Match 100.0%; Score 3025; DB 1; Length 564;  
Best Local Similarity 100.0%; Pred. No. 3e-176;  
Matches 564; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MATAERRALGIGFOWLSLATLVILICAGOGRRDGGPACYGFDLYFILDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFOWLSLATLVILICAGOGRRDGGPACYGFDLYFILDKSGSVLHHWN 60  
QY 61 EIYFVQLAHKFTISPOLRMSFVIFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Db 61 EIYFVQLAHKFTISPOLRMSFVIFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
QY 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSERANRSRDGAIVYCVGV 180  
Db 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSERANRSRDGAIVYCVGV 180  
QY 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHLSLKKSCTEILAAEPSTICAGESFQVY 240  
Db 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHLSLKKSCTEILAAEPSTICAGESFQVY 240  
QY 241 VRGNGFRHARNVDRVLCFSKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGKMAALQVSMN 300  
Db 241 VRGNGFRHARNVDRVLCFSKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGKMAALQVSMN 300  
QY 301 DGLSFSSSVIIITTHCSGDSILAIALILFLLLALLLWFWPLCCTVIKEVPPPPAE 360  
Db 301 DGLSFSSSVIIITTHCSGDSILAIALILFLLLALLLWFWPLCCTVIKEVPPPPAE 360

QY 361 ESEEDDDGLPKKWPPTVDASYGGGKIRMEVRWGEKSTEGAKLEKAKNARVKM 420  
Db 361 ESEEDDDGLPKKWPPTVDASYGGGKIRMEVRWGEKSTEGAKLEKAKNARVKM 420  
QY 421 PEQYEFPEPRNLNNMRSSPRKWTSPKGLDALWLLRKGYDRVSMRPPQDGTGR 480  
Db 421 PEQYEFPEPRNLNNMRSSPRKWTSPKGLDALWLLRKGYDRVSMRPPQDGTGR 480  
QY 481 CINFTRVKNQPAKYPLNNAYHTSSPPAPITYTPPPAPHCPPPPPSAPTPIPPSPSTL 540  
Db 481 CINFTRVKNQPAKYPLNNAYHTSSPPAPITYTPPPAPHCPPPPPSAPTPIPPSPSTL 540  
QY 541 PPPQAPPPNRAPPSPRPPPSV 564  
Db 541 PPPQAPPPNRAPPSPRPPPSV 564  
RESULT 3  
US-60-389-987-1823  
; Sequence 1823, Application US/60389987  
; GENERAL INFORMATION:  
; APPLICANT: Ghosh, Soumitra S.  
; APPLICANT: Fahy, Eoin D.  
; APPLICANT: Zhang, Bing  
; APPLICANT: Gibson, Bradford W.  
; APPLICANT: Taylor, Steven W.  
; APPLICANT: Glenn, Gary M.  
; APPLICANT: Warnock, Dale E.  
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME  
; FILE REFERENCE: 660088.465P2  
; CURRENT APPLICATION NUMBER: US/60/389,987  
; CURRENT FILING DATE: 2002-06-17  
; NUMBER OF SEQ ID NOS: 3025  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1823  
; LENGTH: 564  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-60-389-987-1823

Query Match 100.0%; Score 3025; DB 7; Length 564;  
Best Local Similarity 100.0%; Pred. No. 3e-176;  
Matches 564; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MATAERRALGIGFOWLSLATLVILICAGOGRRDGGPACYGFDLYFILDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFOWLSLATLVILICAGOGRRDGGPACYGFDLYFILDKSGSVLHHWN 60  
QY 61 EIYFVQLAHKFTISPOLRMSFVIFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Db 61 EIYFVQLAHKFTISPOLRMSFVIFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
QY 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSERANRSRDGAIVYCVGV 180  
Db 121 HEGFERASEQIYYENROGYRTASVIALTDGELHEDLFFYSERANRSRDGAIVYCVGV 180  
QY 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHLSLKKSCTEILAAEPSTICAGESFQVY 240  
Db 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHLSLKKSCTEILAAEPSTICAGESFQVY 240  
QY 241 VRGNGFRHARNVDRVLCFSKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGKMAALQVSMN 300  
Db 241 VRGNGFRHARNVDRVLCFSKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGKMAALQVSMN 300  
QY 301 DGLSFSSSVIIITTHCSGDSILAIALILFLLLALLLWFWPLCCTVIKEVPPPPAE 360  
Db 301 DGLSFSSSVIIITTHCSGDSILAIALILFLLLALLLWFWPLCCTVIKEVPPPPAE 360  
QY 361 ESEEDDDGLPKKWPPTVDASYGGGKIRMEVRWGEKSTEGAKLEKAKNARVKM 420

```
Db 361 ESEEDDDGLPKKWTVDASYGGRGVGKIRMEVRWGEKSTEGAKLEKAKNARVKM 420
Qy 421 PEQYEFEPENLNMMRRSSPRKWYSPIKGLDALWLLRKGYDRVSVMRPQGDTR 480
Db 421 PEQYEFEPENLNMMRRSSPRKWYSPIKGLDALWLLRKGYDRVSVMRPQGDTR 480
Qy 481 CINFTRVKNQPAKYPLNNAYHTSSPPAPITYTPPPAPHCPPPPPSAPTPIPSPPSTL 540
Db 481 CINFTRVKNQPAKYPLNNAYHTSSPPAPITYTPPPAPHCPPPPPSAPTPIPSPPSTL 540
Qy 541 PPPQAPPNNRAPPSPRPPSPV 564
Db 541 PPPQAPPNNRAPPSPRPPSPV 564

RESULT 4
PCT-US02-08253-194
; Sequence 194, Application PC/TUS0208253
; GENERAL INFORMATION:
; APPLICANT: Carson-Walter, Eleanor
; APPLICANT: St. Croix, Brad
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00179
; CURRENT APPLICATION NUMBER: PCT/US02/08253
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/308,829
; PRIOR FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 359
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 194
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Mus musculus
PCT-US02-08253-194

Query Match 95.9%; Score 2901; DB 1: Length 562;
Best Local Similarity 96.8%; Pred. No. 1e-168;
Matches 538; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

Qy 9 LGIGFOWLSLATLVLCAGOGGREDGGPACYGDFLYFILDKSGSVLHHNNEIYFVEQ 68
Db 7 LGAGLGLCAALVLCAGHGGREDGGPACYGDFLYFILDKSGSVLHHNNEIYFVEQ 66
Qy 69 LAHFISPOLRMSFIVFSTRTTLMKLTEDREQIROGLEELQKVLPGDVTYMHGFERAS 128
Db 67 LAHFISPOLRMSFIVFSTRTTLMKLTEDREQIROGLEELQKVLPGDVTYMHGFERAS 126
Qy 129 EOYYENRQGYRTASVIITDGLHEDLFFYSREANRSRDLGAIVYCVGVKDFNETQL 188
Db 127 EOYYENSQGYRTASVIITDGLHEDLFFYSREANRSRDLGAIVYCVGVKDFNETQL 186
Qy 189 ARIADSKDHVPVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQVVVRNGNFRH 248
Db 187 ARIADSKDHVPVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQVVVRNGNFRH 246
Qy 249 ARNVDRLCSFKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFSS 308
Db 247 ARNVDRLCSFKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFSS 306
Qy 309 SVIITTHCHSDGSLAIALLLFLLALLLWFWPLCCTVIKEVPPPPAAESEEEDDD 368
Db 247 ARNVDRLCSFKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFSS 306
Qy 309 SVIITTHCHSDGSLAIALLLFLLALLLWFWPLCCTVIKEVPPPPAAESEEEDDD 368
Db 307 SVIITTHCHSDGSLAIALLLFLLALLLWFWPLCCTVIKEVPPPPVEESEEEDDD 366
Qy 369 GLPKKKWPTVDASYGGRGVGKIRMEVRWGEKSTEGAKLEKAKNARVKMPEQYEF 428
Db 367 GLPKKKWPTVDASYGGRGVGKIRMEVRWGEKSTEGAKLEKAKNARVKMPEQYEF 426
Qy 429 EPRNLNNMRRSPSPRWKYSPIKGLDALWLLRKGYDRVSVMRPQGDTRCINFTRVK 488
```

```
Db 427 EPRNLNNMRRSPSPRWKYSPIKGLDALWLLRKGYDRVSVMRPQGDTRCINFTRVK 486
Qy 489 NNQPAKYPLNNAYHTSSPPAPITYTPPPAPHCPPPPPSAPTPIPSPPSTLPPPPQAPP 548
Db 487 NSQPAKYPLNNAYHTSSPPAPITYTPPPAPHCPPPPPSAPTPIPSPPSTLPPPPQAPP 546
Qy 549 PNRAPPPSRPPSPV 564
Db 547 PNRAPPPSRPPSPV 562

RESULT 5
PCT-US02-08253-301
; Sequence 301, Application PC/TUS0208253
; GENERAL INFORMATION:
; APPLICANT: Carson-Walter, Eleanor
; APPLICANT: St. Croix, Brad
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00179
; CURRENT APPLICATION NUMBER: PCT/US02/08253
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/308,829
; PRIOR FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 359
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 301
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Mouse
PCT-US02-08253-301

Query Match 95.9%; Score 2901; DB 1: Length 562;
Best Local Similarity 96.8%; Pred. No. 1e-168;
Matches 538; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

Qy 9 LGIGFOWLSLATLVLCAGOGGREDGGPACYGDFLYFILDKSGSVLHHNNEIYFVEQ 68
Db 7 LGAGLGLCAALVLCAGHGGREDGGPACYGDFLYFILDKSGSVLHHNNEIYFVEQ 66
Qy 69 LAHFISPOLRMSFIVFSTRTTLMKLTEDREQIROGLEELQKVLPGDVTYMHGFERAS 128
Db 67 LAHFISPOLRMSFIVFSTRTTLMKLTEDREQIROGLEELQKVLPGDVTYMHGFERAS 126
Qy 129 EOYYENRQGYRTASVIITDGLHEDLFFYSREANRSRDLGAIVYCVGVKDFNETQL 188
Db 127 EOYYENSQGYRTASVIITDGLHEDLFFYSREANRSRDLGAIVYCVGVKDFNETQL 186
Qy 189 ARIADSKDHVPVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQVVVRNGNFRH 248
Db 187 ARIADSKDHVPVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQVVVRNGNFRH 246
Qy 249 ARNVDRLCSFKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFSS 308
Db 247 ARNVDRLCSFKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFSS 306
Qy 309 SVIITTHCHSDGSLAIALLLFLLALLLWFWPLCCTVIKEVPPPPAAESEEEDDD 368
Db 307 SVIITTHCHSDGSLAIALLLFLLALLLWFWPLCCTVIKEVPPPPVEESEEEDDD 366
Qy 369 GLPKKKWPTVDASYGGRGVGKIRMEVRWGEKSTEGAKLEKAKNARVKMPEQYEF 428
Db 367 GLPKKKWPTVDASYGGRGVGKIRMEVRWGEKSTEGAKLEKAKNARVKMPEQYEF 426
Qy 429 EPRNLNNMRRSPSPRWKYSPIKGLDALWLLRKGYDRVSVMRPQGDTRCINFTRVK 488
Db 427 EPRNLNNMRRSPSPRWKYSPIKGLDALWLLRKGYDRVSVMRPQGDTRCINFTRVK 486
```

Fri Aug 9 10:57:05 2002

```
QY 489 NNQAKYPLNNAYHTSGPPAPAPYTPPPPPHCPHPPPPSAPTPIPSPPSTLPPPPQAPP 548
Db 487 NSQAKYPLNNAYHTSGPPAPAPYTPPPPPHCPHPPPPSAPTPIPSPPSTLPPPPQAPP 546
QY 549 PNRAPPPSRPPRPSV 564
Db 547 PNRAPPPSRPPRPSV 562
Query Match 66.7%; Score 2017; DB 6; Length 381;
Best Local Similarity 98.1%; Pred. No. 3.7e-115;
Matches 371; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
US-10-038-307-4
; Sequence 4, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Murine
US-10-038-307-4
```

```
QY 187 QLARIADSKDHVPVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFOVVRGNGF 246
Db 4 QLARIADSKDHVPVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFOVVRGNGF 63
QY 247 RHARNVRLCSFKINDSVTLNEKPFVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFI 306
Db 64 RHARNVRLCSFKINDSVTLNEKPFVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFI 123
QY 307 SSSVITTHCSGDSILAIALLLFLALALALWFWPLCCTVIKEVPPPPAESEED 366
Db 124 SSSVITTHCSGDSILAIALLLFLALALALWFWPLCCTVIKEVPPPPAESEED 183
QY 367 DDGLPKKWPVTDASYGGRGVGKIKRMEVRWGEKSTEGAKLEKAKNARVKMPEQEYE 426
Db 184 DDGLPKKWPVTDASYGGRGVGKIKRMEVRWGEKSTEGAKLEKAKNARVKMPEQEYE 243
QY 427 FEPRNLNNMRRPSSPRKWYSPKIGKLDALWLLRKGYDRVSVNRPQDGTGRGINFTR 486
Db 244 FEPRNLNNMRRPSSPRKWYSPKIGKLDALWLLRKGYDRVSVNRPQDGTGRGINFTR 303
QY 487 VKNQPAKYPLNNAYHTSGPPAPAPYTPPPPPHCPHPPPPSAPTPIPSPPSTLPPPPQQA 546
Db 304 VKNQPAKYPLNNAYHTSGPPAPAPYTPPPPPHCPHPPPPSAPTPIPSPPSTLPPPPQQA 363
QY 547 PPNRAPPPSRPPRPSV 564
Db 364 PPNRAPPPSRPPRPSV 381
```

```
RESULT 7
US-10-038-307-18
; Sequence 18, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
```

```
QY 1 MATAERRALGIGFQWLSLATLVLCAGOGGRRDGGPACYGFGFDLYFILDKSGSVLHHWN 60
Db 1 MATAERRALGIGFQWLSLATLVLCAGOGGRRDGGPACYGFGFDLYFILDKSGSVLHHWN 60
QY 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120
Db 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120
QY 121 HEGFERASEQIYYENROGYRTASVIIATDGHEDLFFYSEREANRSRDLGAIVYCVGV 180
Db 121 HEGFERASEQIYYENROGYRTASVIIATDGHEDLFFYSEREANRSRDLGAIVYCVGV 180
QY 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFOV 240
Db 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFOV 240
QY 241 VRGNGFRHARNVRLCSFKINDSVTLNEKPFVEDTYLLCPAPILKEVGMKAALQVSMN 300
Db 241 VRGNGFRHARNVRLCSFKINDSVTLNEKPFVEDTYLLCPAPILKEVGMKAALQVSMN 300
QY 301 DGLSFISSSVITTHCSGDS 320
Db 301 DGLSFISSSVITTHCSGDS 320
```

```
Query Match 54.9%; Score 1661; DB 6; Length 551;
Best Local Similarity 100.0%; Pred. No. 2.5e-93;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
US-10-038-307-18
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-18
```

```
RESULT 8
US-10-038-307-2
; Sequence 2, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-2
```

```
Query Match 54.5%; Score 1649; DB 6; Length 333;
Best Local Similarity 100.0%; Pred. No. 7.5e-93;
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MATAERRALGIGFQWLSLATLVLCAGOGGRRDGGPACYGFGFDLYFILDKSGSVLHHWN 60
Db 1 MATAERRALGIGFQWLSLATLVLCAGOGGRRDGGPACYGFGFDLYFILDKSGSVLHHWN 60
QY 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120
Db 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120
```

```
QY 121 HEGFERASEQIYYENRQGYRTASVITATDGLHEDLFFYSEREANRDLGAIYVCVG 180
|||||
Db 121 HEGFERASEQIYYENRQGYRTASVITATDGLHEDLFFYSEREANRDLGAIYVCVG 180

QY 181 KDFNETQARIADSKDHVPVNDGFOALQGIHSILKKSCEIILAAEPSTICAGESFQVV 240
|||||
Db 181 KDFNETQARIADSKDHVPVNDGFOALQGIHSILKKSCEIILAAEPSTICAGESFQVV 240

QY 241 VRNGGFRHARNVDRVLCSPKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300
|||||
Db 241 VRNGGFRHARNVDRVLCSPKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300

QY 301 DGLSFSSSVIIITTHCS 318
|||||
Db 301 DGLSFSSSVIIITTHCS 318

RESULT 9
US-10-038-307-24
; Sequence 24, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-24

Query Match 54.5%; Score 1649; DB 6; Length 345;
Best Local Similarity 100.0%; Pred. No. 7.8e-93;
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATARRALGIGFQWLSLATLVLCACGGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60
|||||
Db 1 MATARRALGIGFQWLSLATLVLCACGGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60

QY 61 EIYFVEQLAHKFIQPLRMSFVSTRGTTLMKLTEDREQIRQGLEELQKVLPGGDTYM 120
|||||
Db 61 EIYFVEQLAHKFIQPLRMSFVSTRGTTLMKLTEDREQIRQGLEELQKVLPGGDTYM 120

QY 121 HEGFERASEQIYYENRQGYRTASVITATDGLHEDLFFYSEREANRDLGAIYVCVG 180
|||||
Db 121 HEGFERASEQIYYENRQGYRTASVITATDGLHEDLFFYSEREANRDLGAIYVCVG 180

QY 181 KDFNETQARIADSKDHVPVNDGFOALQGIHSILKKSCEIILAAEPSTICAGESFQVV 240
|||||
Db 181 KDFNETQARIADSKDHVPVNDGFOALQGIHSILKKSCEIILAAEPSTICAGESFQVV 240

QY 241 VRNGGFRHARNVDRVLCSPKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300
|||||
Db 241 VRNGGFRHARNVDRVLCSPKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300

QY 301 DGLSFSSSVIIITTHCS 318
|||||
Db 301 DGLSFSSSVIIITTHCS 318

RESULT 10
US-10-038-307-20
; Sequence 20, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
```

```
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-20

Query Match 54.5%; Score 1649; DB 6; Length 564;
Best Local Similarity 100.0%; Pred. No. 1.4e-92;
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATARRALGIGFQWLSLATLVLCACGGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60
|||||
Db 1 MATARRALGIGFQWLSLATLVLCACGGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60

QY 61 EIYFVEQLAHKFIQPLRMSFVSTRGTTLMKLTEDREQIRQGLEELQKVLPGGDTYM 120
|||||
Db 61 EIYFVEQLAHKFIQPLRMSFVSTRGTTLMKLTEDREQIRQGLEELQKVLPGGDTYM 120

QY 121 HEGFERASEQIYYENRQGYRTASVITATDGLHEDLFFYSEREANRDLGAIYVCVG 180
|||||
Db 121 HEGFERASEQIYYENRQGYRTASVITATDGLHEDLFFYSEREANRDLGAIYVCVG 180

QY 181 KDFNETQARIADSKDHVPVNDGFOALQGIHSILKKSCEIILAAEPSTICAGESFQVV 240
|||||
Db 181 KDFNETQARIADSKDHVPVNDGFOALQGIHSILKKSCEIILAAEPSTICAGESFQVV 240

QY 241 VRNGGFRHARNVDRVLCSPKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300
|||||
Db 241 VRNGGFRHARNVDRVLCSPKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300

QY 301 DGLSFSSSVIIITTHCS 318
|||||
Db 301 DGLSFSSSVIIITTHCS 318

RESULT 11
US-10-038-307-26
; Sequence 26, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-26

Query Match 54.1%; Score 1636; DB 6; Length 328;
Best Local Similarity 100.0%; Pred. No. 4.6e-92;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATARRALGIGFQWLSLATLVLCACGGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60
|||||
Db 1 MATARRALGIGFQWLSLATLVLCACGGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60
```

```

QY 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120
Db 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120
QY 121 HEGFERASEQIYYENRQGYRTASVIALTDGELHEDLFFYSREANRSDLCIAIYVCVGV 180
Db 121 HEGFERASEQIYYENRQGYRTASVIALTDGELHEDLFFYSREANRSDLCIAIYVCVGV 180
QY 181 KDFNETOLARIADSKDHVPVNDGFOALQGIHSLKKSCEILAAEPSTICAGESFQV 240
Db 181 KDFNETOLARIADSKDHVPVNDGFOALQGIHSLKKSCEILAAEPSTICAGESFQV 240
QY 241 VRNGFRHARNVDRVLCSPKINDSVTLNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMN 300
Db 241 VRNGFRHARNVDRVLCSPKINDSVTLNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMN 300
QY 301 DGLSFISSSVITTHCS 316
Db 301 DGLSFISSSVITTHCS 316

```

```

RESULT 12
US-10-038-307-22
; Sequence 22, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-22

```

```

Query Match 54.0%; Score 1634.5; DB 6; Length 342;
Best Local Similarity 97.2%; Pred. No. 5.9e-92;
Matches 318; Conservative 0; Mismatches 0; Indels 9; Gaps 1;

```

```

QY 1 MATARRALGIGFQWLSLATLVLCAGOGGRRDGGP-----ACYGGFDLYFLDK 51
Db 1 MATARRALGIGFQWLSLATLVLCAGOGGRRDGGPMDYKDDDKACYGGFDLYFLDK 60
QY 52 SGVYLHHWNEIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQK 111
Db 61 SGVYLHHWNEIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQK 120
QY 112 VLPGGDTYMEHGERASEQIYYENRQGYRTASVIALTDGELHEDLFFYSREANRSDLC 171
Db 121 VLPGGDTYMEHGERASEQIYYENRQGYRTASVIALTDGELHEDLFFYSREANRSDLC 180
QY 172 GAIYVCVGVKDFNETOLARIADSKDHVPVNDGFOALQGIHSLKKSCEILAAEPSTI 231
Db 181 GAIYVCVGVKDFNETOLARIADSKDHVPVNDGFOALQGIHSLKKSCEILAAEPSTI 240
QY 232 CAGESFQVVRNGFRHARNVDRVLCSPKINDSVTLNEKPFSEVEDTYLLCPAPILKEVGM 291
Db 241 CAGESFQVVRNGFRHARNVDRVLCSPKINDSVTLNEKPFSEVEDTYLLCPAPILKEVGM 300
QY 292 KAALQVSMNDGLSFISSSVITTHCS 318
Db 301 KAALQVSMNDGLSFISSSVITTHCS 327

```

```

RESULT 13
US-10-038-307-14

```

```

; Sequence 14, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-14

```

```

Query Match 53.9%; Score 1629; DB 6; Length 543;
Best Local Similarity 98.8%; Pred. No. 2.2e-91;
Matches 317; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

```

```

QY 1 MATARRALGIGFQWLSLATLVLCAGOGGRRDGGPACYGGFDLYFLDKSGSVLHHWN 60
Db 1 MATARRALGIGFQWLSLATLVLCAGOGGRRDGGPACYGGFDLYFLDKSGSVLHHWN 60
QY 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120
Db 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120
QY 121 HEGFERASEQIYYENRQGYRTASVIALTDGELHEDLFFYSREANRSDLCIAIYVCVGV 180
Db 121 HEGFERASEQIYYENRQGYRTASVIALTDGELHEDLFFYS--EANRSDLCIAIYVCVGV 178
QY 181 KDFNETOLARIADSKDHVPVNDGFOALQGIHSLKKSCEILAAEPSTICAGESFQV 240
Db 179 KDFNETOLARIADSKDHVPVNDGFOALQGIHSLKKSCEILAAEPSTICAGESFQV 238
QY 241 VRNGFRHARNVDRVLCSPKINDSVTLNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMN 300
Db 239 VRNGFRHARNVDRVLCSPKINDSVTLNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMN 298
QY 301 DGLSFISSSVITTHCS 321
Db 299 DGLSFISSSVITTHCS 319

```

```

RESULT 14
US-10-038-307-16
; Sequence 16, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-16

```

```

Query Match 53.9%; Score 1629; DB 6; Length 543;
Best Local Similarity 98.8%; Pred. No. 2.2e-91;
Matches 317; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

```



Search completed: August 9, 2002, 10:35:57  
Job time: 355 sec

```
Qy 1 MATAERRALGIGFOWLSLATLVLCAGOGGREDGGPACYGFDLYFILDKSGSVLHHWN 60
Db 1 MATAERRALGIGFOWLSLATLVLCAGOGGREDGGPACYGFDLYFILDKSGSVLHHWN 60
Qy 61 EIYFVEQLAHKFISPOLRMSFIVFSTGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120
Db 61 EIYFVEQLAHKFISPOLRMSFIVFSTGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120
Qy 121 HEGFERASEQIYYENRQGYRTASVTIATLDGELHEDLFFYSEREANRSDLGAIYVCVGV 180
Db 121 HEGFERASEQIYYENRQGYRTASVTIATLDGELHEDLFFYSEREANRSDLGAIYVCVGV 178
Qy 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHLSILKKSCIEILAAEPSTICAGESFQVW 240
Db 179 KDFNETQLARIADSKDHVPVNDGFOALQGIHLSILKKSCIEILAAEPSTICAGESFQVW 238
Qy 241 VRGNGFRHARNVDRVLCSEKINDSVTLNEKPFVSEDTYLLCPAPILKEVGMKAALQVSMN 300
Db 239 VRGNGFRHARNVDRVLCSEKINDSVTLNEKPFVSEDTYLLCPAPILKEVGMKAALQVSMN 298
Qy 301 DGLSFISSSVIITTHCSDGS 321
Db 299 DGLSFISSSVIITTHCSPKS 319
```

```
RESULT 15
US-10-038-307-10
; Sequence 10, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038.307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-10
```

Query Match 53.5%; Score 1619; DB 6; Length 543;  
Best Local Similarity 98.4%; Pred. No. 8.8e-91;  
Matches 316; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

```
Qy 1 MATAERRALGIGFOWLSLATLVLCAGOGGREDGGPACYGFDLYFILDKSGSVLHHWN 60
Db 1 MATAERRALGIGFOWLSLATLVLCAGOGGREDGGPACYGFDLYFILDKSGSVLHHWN 60
Qy 61 EIYFVEQLAHKFISPOLRMSFIVFSTGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120
Db 61 EIYFVEQLAHKFISPOLRMSFIVFSTGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120
Qy 121 HEGFERASEQIYYENRQGYRTASVTIATLDGELHEDLFFYSEREANRSDLGAIYVCVGV 180
Db 121 HEGFERASEQIYYENRQGYRTASVTIATLDGELHEDLFFYSEREANRSDLGAIYVCVGV 178
Qy 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHLSILKKSCIEILAAEPSTICAGESFQVW 240
Db 179 KDFNETQLARIADSKDHVPVNDGFOALQGIHLSILKKSCIEILAAEPSTICAGESFQVW 238
Qy 241 VRGNGFRHARNVDRVLCSEKINDSVTLNEKPFVSEDTYLLCPAPILKEVGMKAALQVSMN 300
Db 239 VRGNGFRHARNVDRVLCSEKINDSVTLNEKPFVSEDTYLLCPAPILKEVGMKAALQVSMN 298
Qy 301 DGLSFISSSVIITTHCSDGS 321
Db 299 DGLSFISSSVIITTHSSPKS 319
```

us-09-970-076-6.rapn

Fri Aug 9 10:57:05 2002

-----

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:44:30 ; Search time 507.33 Seconds  
(without alignments)  
391.298 Million cell updates/sec

Title: US-09-970-076-6  
Perfect score: 3025  
Sequence: 1 MATAERRALGIGFQWLSLAT.....QAPPPNRRPPRRPPPSV 564

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/paa/US05\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/paa/US08\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/paa/US080\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep.\*  
7: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep.\*  
8: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep.\*  
9: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep.\*  
10: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep.\*  
11: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep.\*  
12: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep.\*  
13: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep.\*  
14: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep.\*  
15: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep.\*  
16: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep.\*  
17: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep.\*  
18: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep.\*  
19: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep.\*  
20: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep.\*  
21: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep.\*  
22: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep.\*  
23: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep.\*  
24: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep.\*  
25: /cgn2\_6/ptodata/2/paa/US100\_COMB.pep.\*  
26: /cgn2\_6/ptodata/2/paa/US101\_COMB.pep.\*  
27: /cgn2\_6/ptodata/2/paa/US102\_COMB.pep.\*  
28: /cgn2\_6/ptodata/2/paa/US103\_COMB.pep.\*  
29: /cgn2\_6/ptodata/2/paa/US104\_COMB.pep.\*  
30: /cgn2\_6/ptodata/2/paa/US105\_COMB.pep.\*  
31: /cgn2\_6/ptodata/2/paa/US106\_COMB.pep.\*  
32: /cgn2\_6/ptodata/2/paa/US107\_COMB.pep.\*  
33: /cgn2\_6/ptodata/2/paa/US108\_COMB.pep.\*  
34: /cgn2\_6/ptodata/2/paa/US109\_COMB.pep.\*  
35: /cgn2\_6/ptodata/2/paa/US110\_COMB.pep.\*  
36: /cgn2\_6/ptodata/2/paa/US111\_COMB.pep.\*  
37: /cgn2\_6/ptodata/2/paa/US112\_COMB.pep.\*  
38: /cgn2\_6/ptodata/2/paa/US113\_COMB.pep.\*  
39: /cgn2\_6/ptodata/2/paa/US114\_COMB.pep.\*  
40: /cgn2\_6/ptodata/2/paa/US115\_COMB.pep.\*  
41: /cgn2\_6/ptodata/2/paa/US116\_COMB.pep.\*  
42: /cgn2\_6/ptodata/2/paa/US117\_COMB.pep.\*  
43: /cgn2\_6/ptodata/2/paa/US118\_COMB.pep.\*  
44: /cgn2\_6/ptodata/2/paa/US119\_COMB.pep.\*  
45: /cgn2\_6/ptodata/2/paa/US120\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3025	100.0	564	23	US-09-918-715-187
2	3025	100.0	564	23	Sequence 187, App
3	3025	100.0	564	23	Sequence 232, App
4	2901	95.9	562	23	Sequence 6, Appli
5	2901	95.9	562	23	Sequence 194, App
6	2017	66.7	381	1	Sequence 301, App
7	2017	66.7	381	18	Sequence 57, Appl

8	2017	66.7	381	21	US-09-796-753-24	Sequence 24, Appl
9	1894	62.6	368	23	PCT-09-970-076-2	Sequence 2, Appl
10	1893.5	62.6	403	1	PCT-US00-30045-94	Sequence 94, Appl
11	1893.5	62.6	403	1	PCT-US01-11988-621	Sequence 621, App
12	1893.5	62.6	403	22	US-09-833-245-621	Sequence 621, App
13	1874	62.0	403	1	PCT-US00-30045-125	Sequence 125, App
14	1874	62.0	403	1	PCT-US01-11988-620	Sequence 620, App
15	1874	62.0	403	22	US-09-833-245-620	Sequence 620, App
16	1649	54.5	333	1	PCT-US99-31025-9	Sequence 9, Appli
17	1649	54.5	333	16	US-09-223-546-9	Sequence 9, Appli
18	1649	54.5	333	18	US-09-471-179-9	Sequence 9, Appli
19	1649	54.5	333	19	US-09-599-596-9	Sequence 9, Appli
20	1649	54.5	333	21	US-09-796-753-12	Sequence 12, Appl
21	1649	54.5	333	23	PCT-09-970-076-8	Sequence 8, Appli
22	1565	51.7	301	1	PCT-US99-31025-132	Sequence 132, App
23	1565	51.7	301	18	US-09-471-179-132	Sequence 132, App
24	1514	50.0	306	1	PCT-US99-31025-30	Sequence 30, Appl
25	1514	50.0	306	1	PCT-US99-31025-51	Sequence 51, Appl
26	1514	50.0	306	18	US-09-471-179-30	Sequence 30, Appl
27	1514	50.0	306	18	US-09-471-179-51	Sequence 51, Appl
28	1430	47.3	274	1	PCT-US99-31025-50	Sequence 50, Appl
29	1430	47.3	274	18	US-09-471-179-50	Sequence 50, Appl
30	1412.5	46.7	488	1	PCT-US00-05226-59	Sequence 59, Appl
31	1412.5	46.7	488	19	US-09-516-745-59	Sequence 59, Appl
32	1411.5	46.7	487	1	PCT-US00-05226-107	Sequence 107, App
33	1411.5	46.7	487	19	US-09-516-745-107	Sequence 107, App
34	1411.5	46.7	488	1	PCT-US00-05226-2	Sequence 2, Appli
35	1411.5	46.7	488	19	US-09-516-745-2	Sequence 2, Appli
36	1411.5	46.7	488	21	US-09-796-753-52	Sequence 52, Appl
37	1411.5	46.7	488	26	US-60-318-891-10	Sequence 10, Appl
38	1411.5	46.7	488	26	US-60-318-905-20	Sequence 20, Appl
39	1411.5	46.7	488	26	US-60-322-468-6	Sequence 6, Appli
40	1411.5	46.7	488	26	US-60-322-732-10	Sequence 10, Appl
41	1411.5	46.7	488	26	US-60-322-733-6	Sequence 6, Appli
42	1411.5	46.7	488	26	US-60-322-790-16	Sequence 16, Appl
43	1411.5	46.7	488	26	US-60-323-078-22	Sequence 22, Appl
44	1411.5	46.7	488	26	US-60-324-050-2	Sequence 2, Appli
45	1411.5	46.7	488	26	US-60-324-246-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-09-918-715-187  
; Sequence 187, Application US/09918715  
; GENERAL INFORMATION:  
; APPLICANT: Brad St. Croix  
; APPLICANT: Bert Vogelstein  
; APPLICANT: Kenneth Kinzler  
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
; FILE REFERENCE: 1107 00134  
; CURRENT APPLICATION NUMBER: US/09/918,715  
; PRIOR FILING DATE: 2001-08-01  
; PRIOR APPLICATION NUMBER: 60/222,599  
; PRIOR FILING DATE: 2000-08-02  
; PRIOR APPLICATION NUMBER: 60/224,360  
; PRIOR FILING DATE: 2000-08-11  
; PRIOR APPLICATION NUMBER: 60/282,850  
; PRIOR FILING DATE: 2000-04-11  
; NUMBER OF SEQ ID NOS: 358  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 187  
; LENGTH: 564  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-918-715-187

Query Match 100.0%; Score 3025; DB 23; Length 564;  
Best Local Similarity 100.0%; Pred. No. 1.4e-192;  
Matches 564; Conservative 0; Mismatches 0; Indels 0; Gaps 0;





Fri Aug 9 10:57:05 2002

Db 244 FPEPRNLNNRRSPSPRKWYSPKIGKLDALWLLRKGYDRVSVMRPQDGTGRCINFTR 303  
QY 487 VKNQPAKYPLNNAYHTSSPPADIYTPPPAPHCPPPPSAPTPPIPPSPSTLPPPPQA 546  
Db 304 VKNQPAKYPLNNAYHTSSPPADIYTPPPAPHCPPPPSAPTPPIPPSPSTLPPPPQA 363  
QY 547 PPNRAPPPSPPPRPSV 564  
Db 364 PPNRAPPPSPPPRPSV 381  
RESULT 7  
US-09-471-179-57  
; Sequence 57, Application US/09471179  
; GENERAL INFORMATION:  
; APPLICANT: Holtzman, Douglas  
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM  
; FILE REFERENCE: 7853-173  
; CURRENT APPLICATION NUMBER: US/09/471,179  
; PRIOR FILING DATE: 1999-12-23  
; CURRENT FILING DATE: 09/223,546  
; PRIOR FILING DATE: 1998-12-30  
; NUMBER OF SEQ ID NOS: 135  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 57  
; LENGTH: 381  
; TYPE: PRT  
; ORGANISM: Mouse  
US-09-471-179-57

Query Match 66.7%; Score 2017; DB 18; Length 381;  
Best Local Similarity 98.1%; Pred. No. 1e-125; 4; Indels 0; Gaps 0;  
Matches 371; Conservative  
QY 187 QLARIADSKDHVPVNDGFQALQGIHSLKSCIEILAAEPSTICAGESFQVVRGNGF 246  
Db 4 QLARIADSKDHVPVNDGFQALQGIHSLKSCIEILAAEPSTICAGESFQVVRGNGF 63  
QY 247 RHARNVDRVLCSEKINDSVTLNEKPSFVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFI 306  
Db 64 RHARNVDRVLCSEKINDSVTLNEKPSFVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFI 123  
QY 307 SSSVITTHCSDGSILAIALLLFLLALALLWFWPLCCTVILKEVPPPPAESEED 366  
Db 124 SSSVITTHCSDGSILAIALLLFLLALALLWFWPLCCTVILKEVPPPPAESEED 183  
QY 367 DDGLPKKWPVDASYGGGVGGIKRMEVRWGEKSTEEGAKLEKAKNARVKMPEQYE 426  
Db 184 DDGLPKKWPVDASYGGGVGGIKRMEVRWGEKSTEEGAKLEKAKNARVKMPEQYE 243  
QY 427 FPEPRNLNNRRSPSPRKWYSPKIGKLDALWLLRKGYDRVSVMRPQDGTGRCINFTR 486  
Db 244 FPEPRNLNNRRSPSPRKWYSPKIGKLDALWLLRKGYDRVSVMRPQDGTGRCINFTR 303  
QY 487 VKNQPAKYPLNNAYHTSSPPADIYTPPPAPHCPPPPSAPTPPIPPSPSTLPPPPQA 546  
Db 304 VKNQPAKYPLNNAYHTSSPPADIYTPPPAPHCPPPPSAPTPPIPPSPSTLPPPPQA 363  
QY 547 PPNRAPPPSPPPRPSV 564  
Db 364 PPNRAPPPSPPPRPSV 381

RESULT 8  
US-09-796-753-24  
; Sequence 24, Application US/09796753  
; GENERAL INFORMATION:  
; APPLICANT: McCarthy, Sean A.  
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF  
; FILE REFERENCE: 7853-227-999  
; CURRENT APPLICATION NUMBER: US/09/796,753  
; CURRENT FILING DATE: 2001-03-01

QY 189 ARIADSKDHVPVNDGFQALQGIHSLKSCIEILAAEPSTICAGESFQVVRGNGFRH 248  
Db 187 ARIADSKDHVPVNDGFQALQGIHSLKSCIEILAAEPSTICAGESFQVVRGNGFRH 246  
QY 249 ARNVDRVLCSEKINDSVTLNEKPSFVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFI 308  
Db 247 ARNVDRVLCSEKINDSVTLNEKPSFVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFI 306  
QY 309 SVIITTHCSDGSILAIALLLFLLALALLWFWPLCCTVILKEVPPPPAESEEDDD 368  
Db 307 SVIITTHCSDGSILAIALLLFLLALALLWFWPLCCTVILKEVPPPPAESEEDDD 366  
QY 369 GLPKKWPVDASYGGGVGGIKRMEVRWGEKSTEEGAKLEKAKNARVKMPEQYEFP 428  
Db 367 GLPKKWPVDASYGGGVGGIKRMEVRWGEKSTEEGAKLEKAKNARVKMPEQYEFP 426  
QY 429 EPRNLNNRRSPSPRKWYSPKIGKLDALWLLRKGYDRVSVMRPQDGTGRCINFTRVK 488  
Db 427 EPRNLNNRRSPSPRKWYSPKIGKLDALWLLRKGYDRVSVMRPQDGTGRCINFTRVK 486  
QY 489 NQPAKYPLNNAYHTSSPPADIYTPPPAPHCPPPPSAPTPPIPPSPSTLPPPPQAPP 548  
Db 487 NSQPAKYPLNNAYHTSSPPADIYTPPPAPHCPPPPSAPTPPIPPSPSTLPPPPQAPP 546  
QY 549 PNRAPPPSPPPRPSV 564  
Db 547 PNRAPPPSPPPRPSV 562

RESULT 6  
PCT-US99-31025-57  
; Sequence 57, Application PC/TUS9931025  
; GENERAL INFORMATION:  
; APPLICANT: Millennium Pharmaceuticals, Inc.  
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE: 7853-173-228  
; CURRENT APPLICATION NUMBER: PCT/US99/31025  
; PRIOR FILING DATE: 1999-12-23  
; EARLIER FILING DATE: 09/223,546  
; NUMBER OF SEQ ID NOS: 135  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 57  
; LENGTH: 381  
; TYPE: PRT  
; ORGANISM: Mouse  
PCT-US99-31025-57

Query Match 66.7%; Score 2017; DB 1; Length 381;  
Best Local Similarity 98.1%; Pred. No. 1e-125; 4; Indels 0; Gaps 0;  
Matches 371; Conservative  
QY 187 QLARIADSKDHVPVNDGFQALQGIHSLKSCIEILAAEPSTICAGESFQVVRGNGF 246  
Db 4 QLARIADSKDHVPVNDGFQALQGIHSLKSCIEILAAEPSTICAGESFQVVRGNGF 63  
QY 247 RHARNVDRVLCSEKINDSVTLNEKPSFVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFI 306  
Db 64 RHARNVDRVLCSEKINDSVTLNEKPSFVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFI 123  
QY 307 SSSVITTHCSDGSILAIALLLFLLALALLWFWPLCCTVILKEVPPPPAESEED 366  
Db 124 SSSVITTHCSDGSILAIALLLFLLALALLWFWPLCCTVILKEVPPPPAESEED 183  
QY 367 DDGLPKKWPVDASYGGGVGGIKRMEVRWGEKSTEEGAKLEKAKNARVKMPEQYE 426  
Db 184 DDGLPKKWPVDASYGGGVGGIKRMEVRWGEKSTEEGAKLEKAKNARVKMPEQYE 243  
QY 427 FPEPRNLNNRRSPSPRKWYSPKIGKLDALWLLRKGYDRVSVMRPQDGTGRCINFTR 486

;; PRIOR APPLICATION NUMBER: 09/183,175  
;; PRIOR FILING DATE: 1998-10-30  
;; PRIOR APPLICATION NUMBER: 09/223,094  
;; PRIOR FILING DATE: 1998-12-30  
;; PRIOR APPLICATION NUMBER: 09/223,546  
;; PRIOR FILING DATE: 1998-12-30  
;; PRIOR APPLICATION NUMBER: 09/224,246  
;; PRIOR FILING DATE: 1998-12-30  
;; PRIOR APPLICATION NUMBER: 09/259,388  
;; PRIOR FILING DATE: 1999-02-26  
;; PRIOR APPLICATION NUMBER: 09/336,536  
;; PRIOR FILING DATE: 1999-06-18  
;; PRIOR APPLICATION NUMBER: 09/342,687  
;; PRIOR FILING DATE: 1999-06-29  
;; PRIOR APPLICATION NUMBER: 09/345,464  
;; PRIOR FILING DATE: 1999-06-30  
;; PRIOR APPLICATION NUMBER: 09/365,164  
;; PRIOR FILING DATE: 1999-07-30  
;; PRIOR APPLICATION NUMBER: 09/399,723  
;; PRIOR FILING DATE: 1999-09-20  
;; PRIOR APPLICATION NUMBER: 09/409,634  
;; PRIOR FILING DATE: 1999-09-30  
;; PRIOR APPLICATION NUMBER: 09/471,179  
;; PRIOR FILING DATE: 1999-12-23  
;; PRIOR APPLICATION NUMBER: 09/474,071  
;; PRIOR FILING DATE: 1999-12-29  
;; PRIOR APPLICATION NUMBER: 09/474,072  
;; PRIOR FILING DATE: 1999-12-29  
;; PRIOR APPLICATION NUMBER: 09/514,010  
;; PRIOR FILING DATE: 2000-02-25  
;; PRIOR APPLICATION NUMBER: 09/516,745  
;; PRIOR FILING DATE: 2000-03-01  
;; PRIOR APPLICATION NUMBER: 09/572,002  
;; PRIOR FILING DATE: 2000-05-14  
;; PRIOR APPLICATION NUMBER: 09/597,993  
;; PRIOR FILING DATE: 2000-06-19  
;; PRIOR APPLICATION NUMBER: 09/599,596  
;; PRIOR FILING DATE: 2000-06-22  
;; PRIOR APPLICATION NUMBER: 09/630,334  
;; PRIOR FILING DATE: 2000-07-31  
;; PRIOR APPLICATION NUMBER: 09/606,565  
;; PRIOR FILING DATE: 2000-06-29  
;; PRIOR APPLICATION NUMBER: 09/606,317  
;; PRIOR FILING DATE: 2000-06-29  
;; PRIOR APPLICATION NUMBER: 09/665,666  
;; PRIOR FILING DATE: 2000-09-20  
;; PRIOR APPLICATION NUMBER: 09/677,751  
;; PRIOR FILING DATE: 2000-09-30  
;; NUMBER OF SEQ ID NOS: 162  
;; SEQ ID NO 24  
;; LENGTH: 381  
;; TYPE: PRT  
;; ORGANISM: Mouse  
US-09-796-753-24

Query Match 66.7%; Score 2017; DB 21; Length 381;  
Best Local Similarity 98.1%; Pred. No. 1e-125;  
Matches 371; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 187 QIARADSKDHVPVNDGFGAQLGQIIHSILKSCIEILAAEPSTICAGESFOVVVRGNCF 246  
DB 4 QIARADSKDHVPVNDGFGAQLGQIIHSILKSCIEILAAEPSTICAGESFOVVVRGNCF 63  
QY 247 RHARNVDRVLCFKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMNDGLSFI 306  
DB 64 RHARNVDRVLCFKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMNDGLSFI 123  
QY 307 SSSVITTTTHCSDGSLAIALLLFLLALLWFWPLCCCTVIKEVPPPAEERSEED 366

Db 124 SSSVITTTTHCSDGSLAIALLLFLLALLWFWPLCCCTVIKEVPPPAEERSEED 183  
QY 367 DDGLPKKKWPTVDASYGGRGVGGIKRMEVRWGEKSGTEEGAKLEKAKNARVKMPEOYE 426  
Db 184 DDGLPKKKWPTVDASYGGRGVGGIKRMEVRWGEKSGTEEGAKLEKAKNARVKMPEOYE 243  
QY 427 FPEPRNLNNMRRSPSPRKWYSPIKGLDALWLLRKGYDRVSVMRPQDGTGRCINFT 486  
Db 244 FPEPRNLNNMRRSPSPRKWYSPIKGLDALWLLRKGYDRVSVMRPQDGTGRCINFT 303  
QY 487 VKNQPAKYIPLNNAYHTSSPPAPIYTPPPAPHCPPPPPSAPTPTIPSPSTLPPPPQA 546  
Db 304 VKNQPAKYIPLNNAYHTSSPPAPIYTPPPAPHCPPPPPSAPTPTIPSPSTLPPPPQA 363  
QY 547 PPNRAPPPSPRPPPSV 564  
Db 364 PPNRAPPPSPRPPPSV 381  
RESULT 9  
US-09-970-076-2  
;; Sequence 2, Application US/09970076  
;; GENERAL INFORMATION:  
;; APPLICANT: Young, John A.T.  
;; APPLICANT: Bradley, Kenneth A.  
;; APPLICANT: Collier, Robert J.  
;; APPLICANT: Mogridge, Jeremy S.  
;; TITLE OF INVENTION: Anthrax Toxin Receptor  
;; FILE REFERENCE: 960296.97745  
;; CURRENT APPLICATION NUMBER: US/09/970,076  
;; CURRENT FILING DATE: 2001-10-03  
;; PRIOR APPLICATION NUMBER: 60/251,481  
;; PRIOR FILING DATE: 2000-12-05  
;; NUMBER OF SEQ ID NOS: 10  
;; SOFTWARE: Patent In Ver. 2.1  
;; SEQ ID NO 2  
;; LENGTH: 368  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-970-076-2

Query Match 62.6%; Score 1894; DB 23; Length 368;  
Best Local Similarity 100.0%; Pred. No. 1.5e-117;  
Matches 364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MATAERRALGIGFQWLSLATLVLCAGQGRREDGGPACYGFDLYFILDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFQWLSLATLVLCAGQGRREDGGPACYGFDLYFILDKSGSVLHHWN 60  
QY 61 EIYFVEQLAHKFISPOLRMSFIVFSTRTTLMKLTEDREQIROGLEELQKVLPGDVTM 120  
Db 61 EIYFVEQLAHKFISPOLRMSFIVFSTRTTLMKLTEDREQIROGLEELQKVLPGDVTM 120  
QY 121 HEGFERASEOIYYENRQGYRTASVIATLDGELHEDLFFYSERANRSDLAGAIVYCVGV 180  
Db 121 HEGFERASEOIYYENRQGYRTASVIATLDGELHEDLFFYSERANRSDLAGAIVYCVGV 180  
QY 181 KDFNETQLARIADSKDHVPVNDGFGAQLGQIIHSILKSCIEILAAEPSTICAGESFOVV 240  
Db 181 KDFNETQLARIADSKDHVPVNDGFGAQLGQIIHSILKSCIEILAAEPSTICAGESFOVV 240  
QY 241 VRGNFRHARNVDRVLCFSKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSN 300  
Db 241 VRGNFRHARNVDRVLCFSKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSN 300  
QY 301 DGLSFSSSVIITTHCSDGSLAIALLLFLLALLWFWPLCCCTVIKEVPPPAE 360  
Db 301 DGLSFSSSVIITTHCSDGSLAIALLLFLLALLWFWPLCCCTVIKEVPPPAE 360  
QY 361 ESEF 364  
|||||

```

; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 621
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-11988-621

Query Match      62.6%; Score 1893.5; DB 1; Length 403;
Best Local Similarity 94.8%; Pred. No. 1.9e-117;
Matches 367; Conservative 2; Mismatches 7; Indels 11; Gaps 1;

QY 1 MATAERRALGIGFQWLSTLTLVLCAGOGGREDGGPACYGDFLYFILDKSGSVLHHWN 60
Db 1 MATAERRALGIGFQWLSTLTLVLCAGOGGREDGGPACYGDFLYFILDKSGSVLHHWN 60
QY 61 EIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQIRQGLELELQKVLPGGDTYM 120
Db 61 EIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQIRQGLELELQKVLPGGDTYM 120
QY 121 HEGFERASEQIYYENRQGYRTASVIIALTDGELHEDLFFYSERANRSDLGAIYVCVG 180
Db 121 HEGFERASEQIYYENRQGYRTASVIIALTDGELHEDLFFYSERANRSDLGAIYVCVG 180
QY 181 KDFNETQLARIADSKDHVPVNDGFOALOGIIHSILKSKCIELAAEPSTICAGESFQV 240
Db 181 KDFNETQLARIADSKDHVPVNDGFOALOGIIHSILKSKCIELAAEPSTICAGESFQV 240
QY 241 VRNGFPHARNVDRVLCSEKINDSVTLNEKPFVSDVETLLCPAPILKEVGMKAALQVSMN 300
Db 241 VRNGFPHARNVDRVLCSEKINDSVTLNEKPFVSDVETLLCPAPILKEVGMKAALQVSMN 300
QY 301 DGLSFSSSVIIITTHCSGDSILAIALLLIFLLALALLWFWPLCCTVILKEVPPPPAE 360
Db 301 DGLSFSSSVIIITTHCSGDSILAIALLLIFLLALALLWFWPLCCTVILKEVPPPPAE 360
QY 361 ESEEDDDGLPKKWPVDASYGGRG 387
Db 361 ESEVSDHSRMA-----VGQG 376

RESULT 12
US-09-833-245-621
; Sequence 621, Application US/09833245
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229,358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 621
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-245-621

Query Match      62.6%; Score 1893.5; DB 22; Length 403;
Best Local Similarity 94.8%; Pred. No. 1.9e-117;
Matches 367; Conservative 2; Mismatches 7; Indels 11; Gaps 1;

QY 1 MATAERRALGIGFQWLSTLTLVLCAGOGGREDGGPACYGDFLYFILDKSGSVLHHWN 60
Db 1 MATAERRALGIGFQWLSTLTLVLCAGOGGREDGGPACYGDFLYFILDKSGSVLHHWN 60

; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 621
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-11988-621

Query Match      62.6%; Score 1893.5; DB 1; Length 403;
Best Local Similarity 94.8%; Pred. No. 1.9e-117;
Matches 367; Conservative 2; Mismatches 7; Indels 11; Gaps 1;

QY 1 MATAERRALGIGFQWLSTLTLVLCAGOGGREDGGPACYGDFLYFILDKSGSVLHHWN 60
Db 1 MATAERRALGIGFQWLSTLTLVLCAGOGGREDGGPACYGDFLYFILDKSGSVLHHWN 60
QY 61 EIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQIRQGLELELQKVLPGGDTYM 120
Db 61 EIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQIRQGLELELQKVLPGGDTYM 120
QY 121 HEGFERASEQIYYENRQGYRTASVIIALTDGELHEDLFFYSERANRSDLGAIYVCVG 180
Db 121 HEGFERASEQIYYENRQGYRTASVIIALTDGELHEDLFFYSERANRSDLGAIYVCVG 180
QY 181 KDFNETQLARIADSKDHVPVNDGFOALOGIIHSILKSKCIELAAEPSTICAGESFQV 240
Db 181 KDFNETQLARIADSKDHVPVNDGFOALOGIIHSILKSKCIELAAEPSTICAGESFQV 240
QY 241 VRNGFPHARNVDRVLCSEKINDSVTLNEKPFVSDVETLLCPAPILKEVGMKAALQVSMN 300
Db 241 VRNGFPHARNVDRVLCSEKINDSVTLNEKPFVSDVETLLCPAPILKEVGMKAALQVSMN 300
QY 301 DGLSFSSSVIIITTHCSGDSILAIALLLIFLLALALLWFWPLCCTVILKEVPPPPAE 360
Db 301 DGLSFSSSVIIITTHCSGDSILAIALLLIFLLALALLWFWPLCCTVILKEVPPPPAE 360
QY 361 ESEEDDDGLPKKWPVDASYGGRG 387
Db 361 ESEVSDHSRMA-----VGQG 376

RESULT 11
PCT-US01-11988-621
; Sequence 621, Application PC/TUS0111988
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: PCT/US01/11988
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: 60/229,358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25

```



QY	61	EIYFYVEQIAHKFISPOLRMSFTIVSTRGTTLMKLTEDREQIRQGLEELQKVLPGGDITYM	120
Db	61	EIYFYVEQIAHKFISPOLRMSFTIVSTRGTTLMKLTEDREQIRQGLEELQKVLPGGDITYM	120
QY	121	HGCFERASEOIYVENRGQYRTASVIALTDGELHEDLFFYSERANRSIDLCAIVYCVGV	180
Db	121	HGCFERASEOIYVENRGQYRTASVIALTDGELHEDLFFYSERANRSIDLCAIVYCVGV	180
QY	181	KDFNETQLARIADSKDHVPVNDGFOALQGIHLSILKKSCIEILAAEPSTICAGESFQVV	240
Db	181	KDFNETQLARIADSKDHVPVNDGFOALQGIHLSILKKSCIEILAAEPSTICAGESFQVV	240
QY	241	VRNGRFRHARNDRVILCSKFINDSVTLNKSPSVEDTYLLCPAPILKEVGMKAALQVSNM	300
Db	241	VRNGRFRHARNDRVILCSKFINDSVTLNKSPSVEDTYLLCPAPILKEVGMKAALQVSNM	300
QY	301	DGLSFTSSSVIIITTHCSGSIILAIALLILFLLALLALLWWFPLCCTVIIKEVPPPPAAE	360
Db	301	DGLSFTSSSVIIITTHCSGSIILAIALLILFLLALLALLWWFPLCCTVIIKEVPPPPAAE	360
QY	361	ESEEEODDGLPKKKWPTVDASYVGGRG	387
Db	361	ESEVSDHSRMA-----VGGOG	376

```

RESULT 13
PCT-US00-30045-125
; Sequence 125, Application PC/TUS0030045
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: PS708PCT
; CURRENT APPLICATION NUMBER: PCT/US00/30045
; CURRENT FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/163,581
; PRIOR FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: 60/215,133
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 201
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 125
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (175)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (320)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (331)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (368)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US00-30045-125

```

```

Qy 121 HEGFERASEQIYYENRGYRTASVIALTDGELHEDLFFYSEREANRSDLGAIYVCVG 180
Db 121 HEGFERASEQIYYENRGYRTASVIALTDGELHEDLFFYSEREANRSDLGAIYVCVG 180
Qy 181 KDFNETQLARIADSKDHVPVNDGFGQALOGIIHSILKKSCIEILAAEPSTICAGESFQV 240
Db 181 KDFNETQLARIADSKDHVPVNDGFGQALOGIIHSILKKSCIEILAAEPSTICAGESFQV 240
Qy 241 VRGNGFPHARNVDRVLCSFKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSHN 300
Db 241 VRGNGFPHARNVDRVLCSFKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSHN 300
Qy 301 DGLSFSSVSSVITTHCHSDGSILAIALLIIFLLALALLWFWPLCTVTIIKEVPPPPAE 360
Db 301 DGLSFSSVSSVITTHCHSDGSILAIALLIIFLLALALLWFWPLCTVTIIKEVPPPPAE 360
Qy 361 ESEED 366
Db 361 ESEVSD 366

RESULT 14
PCT-US01-11988-620
; Sequence 620, Application PC/TUS0111988
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: PCT/US01/11988
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: 60/229, 358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256, 931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199, 384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 620
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (175)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino a
; NAME/KEY: SITE
; LOCATION: (320)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino a
; NAME/KEY: SITE
; LOCATION: (331)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino a
; NAME/KEY: SITE
; LOCATION: (368)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino a
; PCT-US01-11988-620

```

[illegible]

Db 181 KDFNETQLARIADSKDHVFPVNDGFQALQGIHSLKSKSCIEILAAEPSTICAGESFQVV 240  
QY 241 VRNGFRHARNVDRVLCSEFKINDSVTLNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 KDFNETQLARIADSKDHVFPVNDGFQALQGIHSLKSKSCIEILAAEPSTICAGESFQVV 300  
QY 241 VRNGFRHARNVDRVLCSEFKINDSVTLNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMN 360  
Db 241 VRNGFRHARNVDRVLCSEFKINDSVTLNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMN 360  
QY 301 DGLSFSSSVIIITTHCSDGSILATALLILFLLALALLWFWPLCCTVIIKEVPPPPAE 360  
Db 301 DGLSFSSSVIIITTHCSDGSILATALLILFLLALALLWFWPLCCTVIIKEVPPPPAE 360  
QY 361 ESEED 366  
Db 361 ESEVSD 366

Search completed: August 9, 2002, 10:44:31  
Job time: 869 sec

QY 181 KDFNETQLARIADSKDHVFPVNDGFQALQGIHSLKSKSCIEILAAEPSTICAGESFQVV 240  
Db 181 KDFNETQLARIADSKDHVFPVNDGFQALQGIHSLKSKSCIEILAAEPSTICAGESFQVV 240  
QY 241 VRNGFRHARNVDRVLCSEFKINDSVTLNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRNGFRHARNVDRVLCSEFKINDSVTLNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMN 300  
QY 301 DGLSFSSSVIIITTHCSDGSILATALLILFLLALALLWFWPLCCTVIIKEVPPPPAE 360  
Db 301 DGLSFSSSVIIITTHCSDGSILATALLILFLLALALLWFWPLCCTVIIKEVPPPPAE 360  
QY 361 ESEED 366  
Db 361 ESEVSD 366

RESULT 15  
US-09-833-245-620  
; Sequence 620, Application US/09833245  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: Albumin Fusion Proteins  
; FILE REFERENCE: PF546PCT  
; CURRENT APPLICATION NUMBER: US/09/833,245  
; CURRENT FILING DATE: 2001-04-12  
; PRIOR APPLICATION NUMBER: 60/229, 358  
; PRIOR FILING DATE: 2000-04-12  
; PRIOR APPLICATION NUMBER: 60/256, 931  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/199, 384  
; PRIOR FILING DATE: 2000-04-25  
; NUMBER OF SEQ ID NOS: 2267  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 620  
; LENGTH: 403  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (175)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (320)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (331)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (368)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-833-245-620

Query Match 62.0%; Score 1874; DB 22; Length 403;  
Best Local Similarity 98.6%; Pred. No. 3.7e-116;  
Matches 361; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MATARRALGIGFQWLATLVLCAGQGRREDGGPACYGFDLYFLDKSGSVLHHWN 60  
Db 1 MATARRALGIGFQWLATLVLCAGQGRREDGGPACYGFDLYFLDKSGSVLHHWN 60  
QY 61 EIYVFVQLAHKFISPOLRMSFVSTFRGTTLMKLTEDREQIRQGLEELQKVLPGGDTYM 120  
Db 61 EIYVFVQLAHKFISPOLRMSFVSTFRGTTLMKLTEDREQIRQGLEELQKVLPGGDTYM 120  
QY 121 HEGFERASEQIYYENROGYRTASVIIALTGDELHEDLFFYSERANKSRDLGAIVYCVGV 180  
Db 121 HEGFERASEQIYYENROGYRTASVIIALTGDELHEDLFFYSERANKSRDLGAIVYCVGV 180  
QY 181 KDFNETQLARIADSKDHVFPVNDGFQALQGIHSLKSKSCIEILAAEPSTICAGESFQVV 240  
Db 181 KDFNETQLARIADSKDHVFPVNDGFQALQGIHSLKSKSCIEILAAEPSTICAGESFQVV 240



**THIS PAGE BLANK (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: August 9, 2002, 10:32:11 ; Search time 119.13 Seconds  
(without alignments)  
525.859 Million cell updates/sec

Title: US-09-970-076-6

Perfect score: 3025

Sequence: 1 MATARRALGIGFQWLSLAT.....QAPPPNRPSPRPSPRPSPV 564

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802:\*  
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2017	66.7	381	21 AAB01428	Murine TANGO 197.
2	1893.5	62.6	403	22 AAE01439	Human gene 4 encod
3	1874	62.0	403	22 AAE01469	Human gene 4 encod
4	1649	54.5	333	21 AAB01422	Human TANGO 197.
5	1412.5	46.7	488	21 AAB18456	A human TANGO 216
6	1411.5	46.7	488	21 AAB18447	Amino acid sequenc
7	1409.5	46.6	488	21 AAB18455	A human TANGO 216
8	1408.5	46.6	488	21 AAB18457	A human TANGO 216
9	1397.5	46.2	487	21 AAB18458	A murine TANGO 216
10	1396.5	46.2	487	21 AAB18448	Amino acid sequenc
11	1395.5	46.1	487	21 AAB18460	A murine TANGO 216

12	1393.5	46.1	487	21 AAB18459	A murine TANGO 216
13	1392	46.0	297	22 AAM38976	Human polypeptide
14	1206.5	39.9	587	22 AAU19662	Human novel extrac
15	966	31.9	218	22 AAB92985	Human protein sequ
16	205	6.8	80	22 AA004412	Human polypeptide
17	205	6.8	93	22 AAM84307	Human immune/haema
18	205	6.8	99	22 AA002076	Human polypeptide
19	203.5	6.7	1151	22 ABB61598	Drosophila melanog
20	200	6.6	85	22 AA002036	Human polypeptide
21	200	6.6	85	22 AA004573	Human polypeptide
22	199.5	6.6	615	21 AAG46152	Arabidopsis thalia
23	199.5	6.6	679	21 AAG46151	Arabidopsis thalia
24	199.5	6.6	691	21 AAG46150	Arabidopsis thalia
25	199	6.6	406	22 ABG27250	Novel human diagno
26	199	6.6	470	21 AAG49660	Arabidopsis thalia
27	195	6.4	498	22 AAB73400	Human polypeptel
28	194.5	6.4	88	22 AA002124	Human polypeptide
29	191.5	6.3	306	17 AAR69112	Carrot extensin.
30	191	6.3	325	22 ABG21919	Novel human diagno
31	187	6.2	541	19 AAW37148	Mammalian Eha (Men
32	186	6.1	214	17 AAR69113	Cotton fiber-speci
33	185.5	6.1	439	13 AAR28150	Sugar beet chltna
34	182.5	6.0	502	22 AAM52315	Human WASP protei
35	182.5	6.0	502	22 AAG67333	Amino acid sequenc
36	180	6.0	598	22 ABG14000	Novel human diagno
37	179.5	5.9	446	22 ABB70063	Drosophila melanog
38	179	5.9	527	22 ABB58220	Drosophila melanog
39	178.5	5.9	783	19 AAW37151	Mouse neural Mena+
40	178.5	5.9	787	19 AAW37152	Mouse neural Mena+
41	178.5	5.9	802	19 AAW37153	Mouse neural Mena+
42	178.5	5.9	802	22 AAU09139	Mammalian enabled
43	177.5	5.9	707	22 ABB68526	Drosophila melanog
44	176.5	5.8	520	22 AAM52318	Murine WASP protei
45	176.5	5.8	520	22 AAG67350	Amino acid sequenc

#### ALIGNMENTS

RESULT 1

AAB01428  
ID AAB01428 standard; Protein; 381 AA.

XX AAB01428;

DT 20-OCT-2000 (first entry)

XX Murine TANGO 197.

XX TANGO: 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;  
XX graft versus-host diseases; rheumatoid arthritis; psoriasis;  
XX inflammatory bowel disease; septic shock; ulcerative colitis;  
XX Crohn's disease; chronic myelogenous leukemia; cancer; liver  
XX disease; Hodgkin's disease; osteoarthritis; Lyme's disease;  
XX cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;  
XX systemic lupus erythematosus; transgenic animal; diagnosis;  
XX prognosis; prophylactic; therapeutic; mouse.  
OS Mus musculus.  
XX  
XX WO200039284-A1.  
XX  
XX PD 06-JUL-2000.  
XX  
XX PF 23-DEC-1999; 99WO-US31025.  
XX  
XX PR 30-DEC-1998; 98US-0223546.  
XX  
XX PA (MILL-) MILLENNIUM PHARM INC.  
XX  
XX PI Holtzman DA;  
XX  
XX DR WPI; 2000-465743/40.

Sequence 403 AA;

SECRET 4 encoded secreted protein HWLFR02, SEO ID NO:94.

XX  
KW  
KW  
KW  
KW  
Human; secreted protein; proliferative disorder; cancer; tumour;  
foetal abnormality; developmental abnormality; haematopoietic disorder;



Fri Aug 9 10:57:04 2002

us-09-970-076-6.rag

QY	181	KDFNETOLARIADSKDHVPVNDGFOALQGIHSHILKKSCEITLAAEPSTICAGESFQV	240	CC	disorder associated with aberrant TANGO expression. A wide range
Db	181	kdfnetglariadskdhvfpvndgfgalqgihshilksceitlaaepsticagesfqv	240	XX	of cellular disorders can be treated.
QY	241	VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN	300	SQ	Sequence 333 AA;
Db	241	vrngfrharnvdrvlcsfkindsvtlnekpfsvedtyllcpapilkevgmkaalqvsnn	300		
QY	301	DGLSFSSSVIIITTHCSGSIATLALIFLLALALWFWPLCCTVLIKEVPPPPAE	360		
Db	301	dglfsfssvliitthcsdksilaalilxlalalwfwplcctvlikevppppae	360		
QY	361	ESBED 366			
Db	361	esevsd 366			
RESULT	4				
AAB01422					
ID	AAB01422	standard; Protein: 333 AA.			
XX					
AC	AAB01422;				
XX					
DT	20-OCT-2000	(first entry)			
XX					
DE	Human TANGO 197.				
XX					
KW	TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;				
KW	grat versus-host diseases; rheumatoid arthritis; psoriasis;				
KW	inflammatory bowel disease; septic shock; ulcerative colitis;				
KW	Crohn's disease; chronic myelogenous leukemia; cancer; liver				
KW	disease; Hodgkin's disease; osteoarthritis; Lyme's disease;				
KW	cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;				
KW	systemic lupus erythematosus; transgenic animal; diagnosis;				
KW	prognosis; prophylactic; therapeutic; human.				
XX					
OS	Homo sapiens.				
XX					
PN	WO200039284-A1.				
XX					
PD	06-JUL-2000.				
XX					
PF	23-DEC-1999; 99WO-US31025.				
XX					
PR	30-DEC-1998; 98US-0223546.				
XX					
PA	(MILL-) MILLENNIUM PHARM INC.				
XX					
PI	Holtzman DA;				
XX					
DR	WPI; 2000-465743/40.				
DR	N-PSDB; AAA47455.				
XX					
PT	Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,				
PT	224 and 239 polypeptides useful for the treatment of asthma, rheumatoid				
PT	arthritis, psoriasis and autoimmune diseases				
XX					
PS	Claim 8; Fig 4; 209pp; English.				
XX					
CC	Nucleic acids encoding TANGO polypeptides are useful as modulating				
CC	agents for regulating cellular processes like asthma, graft				
CC	versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory				
CC	bowel disease, septic shock, ulcerative colitis, Crohn's disease,				
CC	chronic myelogenous leukemia, cancer, liver disease, Hodgkin's				
CC	disease, osteoarthritis, Lyme's disease, cachexia and autoimmune				
CC	diseases e.g. myasthenia gravis, autoimmune diabetes and systemic				
CC	lupus erythematosus. The nucleic acids are also useful for producing				
CC	transgenic animals and the TANGO polypeptides themselves. Partial				
CC	TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in				
CC	forensic biology, for diagnostic assays, prognostic assays,				
CC	pharmacogenomics and for monitoring clinical trials. TANGO				
CC	polypeptides are suitable for both prophylactic and therapeutic				
CC	methods for treating a subject at risk of a disorder or having a				

CC	disorder associated with aberrant TANGO expression. A wide range				
XX	of cellular disorders can be treated.				
SQ	Sequence 333 AA;				
Query Match	54.5%;	Score 1649;	DB 21;	Length 333;	
Best Local Similarity	100.0%;	Pred. No. 2.2e-122;			
Matches 318;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	MATARRALGIGFQWLSLATLVLCAGGRRDGGPACYGFDLYFLDKSGSVLHHWN	60		
Db	1	mataerralgigfqlslatlvlicagggrrddgppacygfdlyflldksvglhwn	60		
QY	61	EIYFVEQLAHKFIQPQLRMSFIVSTRGTTLMKLTREQIRQGLEELQKVLPGDITYM	120		
Db	61	eiyfveqlahkfispqlrmsfivstrgttlmkltedreqirggleelqkvlpgditym	120		
QY	121	HEGFERASEQIYYENRQGYRTASVITATLDTGELHEDLFFYSEREANRSRDGLAIVCYGV	180		
Db	121	hegferaseqiyyenrqqyrtasvialtdgelhedlffysereanrdrdgalvycgv	180		
QY	181	KDFNETOLARIADSKDHVPVNDGFOALQGIHSHILKKSCEITLAAEPSTICAGESFQV	240		
Db	181	kdfnetglariadskdhvfpvndgfgalqgihshilksceitlaaepsticagesfqv	240		
QY	241	VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN	300		
Db	241	vrngfrharnvdrvlcsfkindsvtlnekpfsvedtyllcpapilkevgmkaalqvsnn	300		
QY	301	DGLSFSSSVIIITTHCS 318			
Db	301	dglfsfsssvliitthcs 318			
RESULT	5				
AAB18456					
ID	AAB18456	standard; Protein: 488 AA.			
XX					
AC	AAB18456;				
XX					
DT	15-JAN-2001	(first entry)			
XX					
DE	A human TANGO 216 polypeptide clone.				
XX					
KW	TANGO 266; TANGO 216;; TANGO 261; TANGO 262; TANGO 267;				
KW	cellular proliferation; cellular differentiation; cellular adhesion;				
KW	von Willebrand factor-associated disorder; cell trafficking; cancer;				
KW	hematopoietic associated disease; atelectasis; pulmonary congestion;				
KW	oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;				
KW	intestinal disorder; spleen associated disease; renal disorder;				
KW	cardiovascular disorder; ischemic heart disease; hydrocephalus;				
KW	brain herniation; latrogenic disease; inflammation; meningitis;				
KW	Alzheimer's disease; cerebral toxoplasmosis; Parkinson's disease;				
KW	multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.				
OS	Homo sapiens.				
XX					
PN	WO200052022-A1.				
XX					
PD	08-SEP-2000.				
XX					
PF	01-MAR-2000; 2000WO-US05226.				
XX					
PR	01-MAR-1999; 99US-0122458.				
XX					
PA	(MILL-) MILLENNIUM PHARM INC.				
XX					
PI	Barnes TM, Holtzman DA, Sharp JD, Fraser CC;				
XX					
DR	WPI; 2000-579269/54.				
DR	N-PSDB; AAA75158.				
XX					



Novel human and murine secreted proteins designated TANGO 216, 261, 262, 266 and 267 useful as modulating agents of cellular processes, e.g. for treating cancer -

Disclosure: Page -: 175pp; English.

AAB18455-57 represent human TANGO 216 proteins. The specification also describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO polypeptides can be used to modulate cellular proliferation, modulate cellular differentiation and/or modulate cellular adhesion. The proteins can be used to treat any von Willebrand factor-associated disorder, regulate extracellular matrix structuring, cellular adhesion, and cell trafficking and/or migration, modulate cellular interactions, modulate cell adhesion in proliferative disorders, such as cancer, modulate the proliferation, differentiation, and/or function of cells that appear in the bone marrow, and leukocytes, treat bone marrow, blood and hematopoietic associated diseases and disorders, atelectasis, pulmonary congestion or oedema, emphysema, chronic bronchitis, bronchial asthma and bronchiectasis, intestinal disorders, spleen associated diseases, modulate renal disorders, treat cardiovascular disorders such as ischemic heart disease, modulate the proliferation, differentiation, and/or function of bone and cartilage cells and to treat bone and/or cartilage associated diseases or disorder. They may also be used to treat disorders associated with the ovaries, and cerebral oedema, hydrocephalus, brain herniations, iatrogenic disease, inflammations, bacterial and viral meningitis, Alzheimer's Disease, cerebral toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers, hydrocephalus and encephalitis, and treat hepatic disorders.

note: the present sequence does not appear in the specification; it was created using information provided.

Sequence 488 AA:

Query Match 46.78; Score 1412.5; DB 21; Length 488;  
 Best Local Similarity 55.66; Pred. No. 2e-103;  
 Matches 274; Conservative 77; Mismatches 123; Indels 19; Gaps 4;

Qy 4 AERRALGIGFOWLSLATLVLCAGGG--RRDDGGPACVGGFDLYFLDKSGSVLHHWNEI 62  
 Db 3 aersparspgswlfpglwlvlspgglrraqdpcscrafdlyfvdksqsvannwiel 62  
 Qy 63 YFVEQLAHKFIISPOLRMSFTVFSTRGTTLMKLTEDREIQGLEELQKVLPGDGYMHE 122  
 Db 63 ynfvgqlaerfvspemlsflvfssqatllptlgtgdkiskgledlkrvspgetyih 122  
 Qy 123 GFERASQIYYENRQGYRTASVIALTDGELHEDLFFYSEREANRDLGAIVCVGVKD 182  
 Db 123 glklaneqi--qkagglktssliialtdgldgvlpsyaekaisrslgasvycvgvld 180  
 Qy 183 FNETQLARIADSKDHFVPVNDGFOALQGIHSILKSCIEILAAEPSPICAGESFOVVVR 242  
 Db 181 feqqleriadskeqfvpkvggfgalqglnsilaqstcellelqpsvsvgeefqlvis 240  
 Qy 243 GNGFRHARNVDRVLCSFKINDSVFLNKPFSVEDTYLILCAPILKEVGMKAALQVSMNDG 302  
 Db 241 grgfmlgrngsvlctyvtvntytttsvkpsvqnsmlcpapilnkagetldvsvsfng 300  
 Qy 303 LSFSSSVIITTHCSDGSIILAILLIFLLALALLWFWPLCLCTVVIKEVPPPPABES 362  
 Db 301 ksvlsgslivtatecsngiaailvlllllglglmwfwplcckvkvkpppppppap 360  
 Qy 363 EEDDGLPKKKWPTVDASYGGGVGGIKRMEVRWKGSTEGAKLEKAKNARVKMPE 422  
 Db 361 keeseelpthkwtvdasyggvggikmevrwgdkgstegarlekaknavvkipe 420  
 Qy 423 QEYEFPEPRNLNNNRSSPR-----KWYGPICKGLDALVLLRKGYDRVSMRPP 474  
 Db 421 eteeiprp-----rprkphqpqtkwytpikgrldalwallrrqydrvslmfpq 472  
 Qy 475 PGDTGRCINFTRY 487  
 Db 473 egdegrcinfsvr 485

RESULT 6  
 AAB18447  
 ID AAB18447 standard; Protein; 488 AA.  
 XX  
 AC AAB18447;  
 XX  
 DT 15-JAN-2001 (first entry)  
 XX  
 DE Amino acid sequence of human TANGO 216 polypeptide.  
 XX  
 KW TANGO 266; TANGO 216;; TANGO 261; TANGO 262; TANGO 267;  
 KW cellular proliferation; cellular differentiation; cellular adhesion;  
 KW von Willebrand factor-associated disorder; cell trafficking; cancer;  
 KW hematopoietic associated disease; atelectasis; pulmonary congestion;  
 KW oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;  
 KW intestinal disorder; spleen associated disease; renal disorder;  
 KW cardiovascular disorder; ischemic heart disease; hydrocephalus;  
 KW brain herniation; iatrogenic disease; inflammation; meningitis;  
 KW Alzheimer's Disease; cerebral toxoplasmosis; Parkinson's disease;  
 KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..34  
 FT Domain /note= "signal peptide"  
 FT Protein /note= "extracellular domain"  
 FT Domain /note= "mature protein"  
 FT Domain /note= "von Willebrand factor A domain"  
 FT Domain /note= "transmembrane domain"  
 FT Domain /note= "cytoplasmic domain"  
 FT Domain /note= "transmembrane domain"  
 FT Domain /note= "extracellular domain"  
 WO2000052022-A1.  
 PD 08-SEP-2000.  
 XX 01-MAR-2000; 2000WO-US05226.  
 XX 01-MAR-1999; 99US-0122458.  
 (MILL-) MILLENNIUM PHARM INC.  
 PI Barnes TM, Holtzman DA, Sharp JD, Fraser CC;  
 WPI: 2000-579269/54.  
 N-PSDB; AAA75149.  
 XX Novel human and murine secreted proteins designated TANGO 216, 261, 262, 266 and 267 useful as modulating agents of cellular processes, e.g. for treating cancer -  
 XX Claim 8; Fig 1A-C; 175pp; English.  
 CC The present sequence represents a human TANGO 216 polypeptide. The specification also describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO polypeptides can be used to modulate cellular proliferation, modulate cellular differentiation and/or modulate cellular adhesion. The proteins can be used to treat any von Willebrand factor-associated disorder, regulate extracellular matrix structuring, cellular adhesion, and cell trafficking and/or migration, modulate cellular interactions, modulate cell adhesion in proliferative disorders, such as cancer, modulate the proliferation, differentiation,

Fri Aug 9 10:57:04 2002

CC and/or function of cells that appear in the bone marrow, and leukocytes,  
CC treat bone marrow, blood and hematopoietic associated diseases and  
CC disorders, atelectasis, pulmonary congestion or oedema, emphysema,  
CC chronic bronchitis, bronchial asthma and bronchiectasis, intestinal  
CC disorders, spleen associated diseases, modulate renal disorders, treat  
CC cardiovascular disorders such as ischemic heart disease, modulate the  
CC proliferation, differentiation, and/or function of bone and cartilage  
CC cells and to treat bone and/or cartilage associated diseases or  
CC disorder. They may also be used to treat disorders associated with the  
CC ovaries, cerebral oedema, hydrocephalus, brain herniations, iatrogenic  
CC disease, inflammations, bacterial and viral meningitis, Alzheimer's  
CC disease, cerebral toxoplasmosis, Parkinson's disease, multiple sclerosis,  
CC brain cancers, hydrocephalus and encephalitis, and treat hepatic  
CC disorders.  
XX  
SQ Sequence 488 AA;

Query Match 46.7%; Score 1411.5; DB 21; Length 488;  
Best Local Similarity 55.6%; Pred. No. 2.3e-103;  
Matches 274; Conservative 77; Mismatches 123; Indels 19; Gaps 4;

QY 4 AERRALGIFQWLSLATLVLCAGQGG-RRDGGPACYGGFDLYFILDKSGSVLHWHNEI 62  
DB 3 aersparspgswlfpglwllvsgpggllraqeqpsrrafldlyfvdksgvannwiel 62  
QY 63 YVFVQLAHKFIQPLRMSFTVSTRTTLMKLTEDREQIQLGEEQLKVLPGGDYMH 122  
DB 63 ynfvgqlaerspmlrslfvsfsqatllptlgtgdkiskgledlkrvspvgetyih 122  
QY 123 GFERASEQIYENRQYRTASVIITALDGHEDLFFYSEREANRSLDGAIVYCVGVD 182  
DB 123 gliklaneqi--qkagglktssilialtdgldglvpsyaekaisrslgasvycvgvld 180  
QY 183 FNETQARIANDKHVPVNDGFOALQGIHSILKSKSIEILAAEPSTICAGESFQVVR 242  
DB 181 feaqglerladskeqfvpkggfgalkginsilaqstlellqpssvcvgeefqivls 240  
QY 243 GNGFHRARNDVLGSFKINDSVTLNKPFSVEDTYLLCPAPILKEVGMKAALOVSN 302  
DB 241 grgfmlgrngsvlctyvtvnetysvkpsvqlnsmicpalknagetldvsvsfng 300  
QY 303 LFSSIVITTHSGSILAIALLLFLALALLWFWPLCCTVIIEVPPAPAES 362  
DB 301 ksvigsllvtatecsnglaaiivilvlllglgmwfwplcckvklpdpptpppap 360  
QY 363 EEDDDGLPKKWTVDASYGGRGVGGIKRMEVRWKGSTEECAKLEKAKNARVKPE 422  
DB 361 keeeepktpkwtvdasyggvggikrmevrwgdkgsteegarlekaknavvkipe 420  
QY 423 QEYFEPPEPNLNNMRPSSPR-----KWYSPITKGLDALWLLRRKGYDRVSMRPQ 474  
DB 421 eteeplr-----rpprpthppqtkwypikgrldalwallrrqydrvsmlrpq 472  
QY 475 PGDTGRCINFRV 487  
DB 473 egdegrcinfsv 485

RESULT 7  
AA18455  
ID AAB18455 standard; Protein; 488 AA.  
XX  
AC AAB18455;  
XX  
DT 15-JAN-2001 (first entry)  
XX  
DE A human TANGO 216 polypeptide clone.  
XX  
KW TANGO 266; TANGO 216; TANGO 262; TANGO 267;  
KW cellular proliferation; cellular differentiation; cellular adhesion;  
KW von Willebrand factor-associated disorder; cell trafficking; cancer;  
KW hematopoietic associated disease; atelectasis; pulmonary congestion;

KW oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;  
KW intestinal disorder; spleen associated disease; renal disorder;  
KW cardiovascular disorder; ischemic heart disease; hydrocephalus;  
KW brain herniation; iatrogenic disease; inflammation; meningitis;  
KW Alzheimer's Disease; cerebral toxoplasmosis; Parkinson's disease;  
KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.  
XX Homo sapiens.  
OS  
XX WO200052022-A1.  
PN  
XX  
XX  
PD 08-SEP-2000.  
XX  
XX 01-MAR-2000; 2000WO-US05226.  
XX  
XX 01-MAR-1999; 99US-0122458.  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
XX  
XX Barnes TM, Holtzman DA, Sharp JD, Fraser CC;  
XX  
XX WPI; 2000-579269/54.  
XX N-PSDB; AAA75157.  
XX  
XX Novel human and murine secreted proteins designated TANGO 216, 261,  
XX 262, 266 and 267 useful as modulating agents of cellular processes,  
XX e.g. for treating cancer -  
XX  
XX Disclosure; Page -; 175pp; English.  
XX  
XX AAB18455-57 represent human TANGO 216 proteins. The specification also  
XX describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO  
XX polypeptides can be used to modulate cellular proliferation, modulate  
XX cellular differentiation and/or modulate cellular adhesion. The  
XX proteins can be used to treat any von Willebrand factor-associated  
XX disorder, regulate extracellular matrix structuring, cellular adhesion,  
XX and cell trafficking and/or migration, modulate cellular interactions,  
XX modulate cell adhesion in proliferative disorders, such as cancer,  
XX modulate the proliferation, differentiation, and/or function of cells  
XX that appear in the bone marrow, and leukocytes, treat bone marrow, blood  
XX and hematopoietic associated diseases and disorders, atelectasis,  
XX asthma and bronchiectasis, intestinal disorders, chronic bronchitis, bronchial  
XX pulmonary congestion or oedema, emphysema, treat cardiovascular disorders such  
XX as ischemic heart disease, modulate the proliferation, differentiation,  
XX and/or function of bone and cartilage cells and to treat bone and/or  
XX cartilage associated diseases of disorder. They may also be used to  
XX treat disorders associated with the ovaries, and cerebral oedema,  
XX hydrocephalus, brain herniations, Alzheimer's disease, cerebral  
XX bacterial and viral meningitis, Parkinson's disease, multiple sclerosis, brain cancers,  
XX toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers,  
XX hydrocephalus and encephalitis, and treat hepatic disorders.  
XX note: the present sequence does not appear in the specification; it was  
XX created using information provided.

Sequence 488 AA;

Query Match 46.6%; Score 1409.5; DB 21; Length 488;  
Best Local Similarity 56.2%; Pred. No. 3.4e-103;  
Matches 271; Conservative 77; Mismatches 115; Indels 19; Gaps 4;

QY 15 WLSLATLVLCAGQGG-RRDGGPACYGGFDLYFILDKSGSVLHWHNEIYVFPQLAHKF 73  
DB 14 wlfpgllwllvsgpggllraqeqpsrrafldlyfvdksgvannwielynfvqlaerf 73  
QY 74 ISPOLRMSFTVSTRTTLMKLTEDREQIQLGEEQLKVLPGGDYMHGFERASEQIYY 133  
DB 74 vspemrlsfvssqatllptlgtgkiskgledlkrvspvgetyihgliklaneqi-- 131  
QY 134 ENRQGYRTASVIITALDGHEDLFFYSEREANRSLDGAIVYCVGVDNETQLARIAD 193  
DB 132 qkagglktssilialtdgldglvpsyaekaisrslgasvycvgvldfeaqdleriad 191







Db 120 ihēglkianeqī --qnagglkassiiiaitdgldglpvsaeneakrsigasyvcgv 177

Qy 180 VKDFNETQLARIADSKDHVPPVNDGFGAQTTHSLKSCIEIIAAESTICAGESFQV 239

Db 178 vldfēgaqlariadskdqfpvkvgfgalkgiinsilagctellelspssvvcgekfgv 237



us-09-970-076-6.rag

Fri Aug 9 10:57:04 2002

```
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0218290.
PR 14-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 26-JUL-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PR XX
PR XX (HUMA-) HUMAN GENOME SCI INC.
PR XX Rosen CA, Barash SC, Ruben SM;
PR XX WPI; 2001-465572/50.
PR XX N-PSDB; RAS31233.
PR XX Nucleic acid molecules encoding human secreted extracellular matrix
PR XX proteins, used in preventing, treating or ameliorating a disorder, e.g.
PR XX Alzheimer's and Parkinson's diseases and cancers.
PR XX
```



PS	Claim 11; SEQ ID No 312; 577pp; English.	
XX	The invention relates to isolated nucleic acid molecules encoding	
CC	novel human secreted extracellular matrix proteins (SPs). The	
CC	polynucleotides and proteins are used to prevent, treat a medical	
CC	condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,	
CC	chickens or sheep. For example, disorders associated with decreased	
CC	expression of SPs. The SP polynucleotide or a vector expressing them may	
CC	be administered to treat diseases by gene therapy. Antisense molecules	
CC	may be administered to down regulate expression of SPs by binding with	
CC	the cells own genes and preventing their expression. The polynucleotides	
CC	may also be used as DNA probes in diagnostic assays. The SPs may also be	
CC	used as antigens to produce antibodies and to identify modulators	
CC	(agonists and antagonists) of the SPs. The anti-(SP) antibodies and	
CC	antagonists may also be used to down regulate expression and activity of	
CC	SP and as diagnostic agents for detecting the presence of SPs in samples.	
CC	The disorders include for example: immune/autoimmune diseases (e.g. HIV	
CC	(human immunodeficiency virus) infections, anaemia, rheumatoid arthritis	
CC	and multiple sclerosis), cancers and hyperproliferative disorders (e.g.	
CC	melanomas, neoplasms of the breast or liver, Sezary syndrome and	
CC	Caucher's disease), neurological diseases (e.g. Alzheimer's disease,	
CC	Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac	
CC	arrest, tachycardia and angina), infections caused by bacteria, viruses	
CC	and fungi and ocular disorders (e.g. corneal infections). Other uses	
CC	include wound healing, maintenance of organs before transplantation,	
CC		
Query Match	39.9%; Score 1206.5; DB 22; Length 587;	
Best Local Similarity	56.2%; Pred. No. 5.1e-87;	
Matches	232; Conservative 67; Mismatches 111; Indels 3; Gaps 2;	
QY	15 WLSLATLVLCAGOGG-RREDGGPACYGFDLYFLDKSGVLHWNHNIYFYVEQLAHKF 73	
DB	170 wlfpglwllvlsppglllraqeqpsccrrafdlvldksgsvanxwlelynfvgqlaerf 229	
QY	74 ISQRLMSFVFSFTRGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFEFASQIYY 133	
DB	230 xspemrlsfiivssqatfilpltdgrgkiskgledkrvspggetyihiegiklaneqi-- 287	
QY	134 ENQGYRTASVITALTDGELHEDLFFYSREANRSRDLGAIVYCVGVKDFNETQLARIAD 193	
DB	288 qkagglktxiialtdgklidglvpsyaekaisrslgasvycvgvldfeqqleriad 347	
QY	194 SKDRHFVNDFOALQGIHSLKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVD 253	
DB	348 skeqfvpvkgfgqalkglnslagctellelqpsvcvgeefivlsrgfmlgsrng 407	
QY	254 RVLCSPKINDSVTLNRPKPSVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFSSSVIIT 313	
DB	408 svlctyvtvnetytskvpsvqnsmlcpapiinkagetldvsvngsgkvsagslivt 467	
QY	314 TTHCSGSILAIALLIFLLALALWFWPLCTVIKEVPPPPPAESEEEDDGLPKK 373	
DB	468 axxcnsnglaaiwilvlllllglglmwfwplcckvixkldppppppppapkeeeelpk 527	
QY	374 KWTVDASYGGRGVGIGKRMVYRWGCKSTEECAKLEKAKNARVKMPEQEYE 426	
DB	528 kwptvdasygygvggikrmvrvdgksgteegarlekaknavvkipteete 580	
RESULT	15	
AAAB92985		
ID	AAAB92985 standard; Protein; 218 AA.	
XX		
AC	AAAB92985;	
XX		
DT	26-JUN-2001 (first entry)	
XX		
DE	Human protein sequence SEQ ID NO:11706.	
XX		
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy.	
XX		
OS	Homo sapiens.	
XX		

us-09-970-076-6.rag

Fri Aug 9 10:57:04 2002

Search completed: August 9, 2002, 10:32:13  
Job time: 132 sec

GenCore version 4.5  
Copyright (C) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:34:15 ; Search time 61.73 Seconds  
(without alignments)  
877.926 Million cell updates/sec

Title: us-09-970-076-6  
Perfect score: 3025  
Sequence: 1 MATAERRALGIGFWLSLAT.....QAPPNRPSPRPSPRPSPV 564

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_71.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	227	7.5	847	2 F96531	hypothetical prote
2	205.5	6.8	760	2 T06291	extensin homolog I
3	205	6.8	494	2 B96534	hypothetical prote
4	203	6.7	599	2 T10798	phenophorin-s - Vo
5	202	6.7	464	2 S22697	extensin - Volvox
6	201	6.6	1201	2 G86441	unknown protein [i
7	200.5	6.6	383	2 AE2295	hypothetical prote
8	200.5	6.6	3124	1 A40020	collagen alpha 1(X
9	199.5	6.6	1392	2 T51947	probable transcrip
10	199	6.6	485	2 A33647	sulfated surface g
11	198	6.5	242	2 S54156	extensin-like prot
12	197.5	6.5	141	2 A34043	hypothetical prote
13	195.5	6.5	699	2 T05225	extensin homolog F
14	194	6.4	727	2 C84534	hypothetical prote
15	191.5	6.3	181	2 S14974	extensin class I (
16	191.5	6.3	306	2 A24354	extensin precursor
17	191.5	6.3	489	2 T11622	extensin class I p
18	191.5	6.3	708	2 D96711	hypothetical prote
19	191	6.3	322	2 S25299	extensin precursor
20	189.5	6.3	388	2 S25298	extensin (clone To
21	189	6.2	368	2 C29356	hydroxyproline-ric
22	188.5	6.2	1206	2 S24407	formin isoform IV
23	188.5	6.2	1468	2 S11515	formin - mouse
24	186.5	6.2	163	2 A29356	hydroxyproline-ric
25	186	6.1	214	2 T10737	extensin-like cell
26	186	6.1	214	2 T09854	proline-rich cell
27	186	6.1	280	2 T11671	extensin-like prot
28	185.5	6.1	439	2 S51939	chitinase (EC 3.2.
29	185.5	6.1	620	2 S06733	hydroxyproline-ric

ALIGNMENTS

RESULT 1

F96531

hypothetical protein F13F21.7 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: F96531

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maizl, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: F96531

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-847 <STO>

A:Cross-references: GB:AE005173; NID:G5430752; PIDN:AAD43152.1; GSPDB:GN00141

C:Genetics:

A:Gene: F13F21.7

A:Map position: 1

Query Match 7.5%; Score 227; DB 2; Length 847;  
Best Local Similarity 31.5%; Pred. No. 3.7e-07;  
Matches 75; Conservative 15; Mismatches 60; Indels 88; Gaps 13;

QY 355 PPPPAEES-----EEEDDGLPKKWTVDASYYGGRVGGIKRMEVRGCKSTEGAKL 410

Db 435 PQQPHESPKPEEPKHELKPKQ-----ESPKPQPSKPEDSPKP 474

QY 411 EKAKNARKVMPEQYEFPEPRLNNMRRPSPRPKW-----YSPIRGKLDALWV 459

Db 475 EQPKPEESPKPQPP-QIPEP-----TKVPSPFPPNAOQPTDDPDYDASPVK----- 518

QY 460 LLRKGYDRVSMRQPGDTRGRCINFRVKNQPAKYLNNAYHTSSPPAPITYTPPPAP 519

Db 519 -----NRRSPPPKVED-----TRVPPQP---PM-----PSPSPSPSIYSPPPV- 556

QY 520 HCPPPPPP-SAPTPP---TPSPPTLPPPPQAPPNRPAPP-----SRPPRP 562

Db 557 HSPPPPVSSPPPPHVYSPPPVSPVSPPPSPPPVHSPPPPVFSPPPVFPSPPPSP 614

RESULT 2

T06291

extensin homolog T9E8.80 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

```

RESULT      4
T10798
pherophorin-s - Volvox carteri
C:Species: Volvox carteri
16-Jul-1999 #sequence revision 16-Jul-1999 #text_change 21-Jul-2000

```

Query Match	6.7%;	Score 202;	DB 2;	Length 464;
Best Local Similarity	23.0%;	Pred. NO. 7.6e-06;		
Matches	100;	Conservative 46;	Mismatches 128;	Indels 160;
Gaps	22;			
QY	208	LOGIIHSILKK-----SCIEIIAAEPST--ICAGESFQVVVRGNGFRHARNVDRVLCSPK	260	
Db	19	LSGIVSOLUSKWLNPVTLRTVYSQPSAAELSTNFAIVYSKGOG-----SYW	66	
QY	261	INDSVTLNEKPSFVEDTYLLCPALIKVEGKMAALQVSMNDGLSFI---SSSVLIITTHC	317	
Db	67	ITEGLSN-----STKVNDLLTFVRNGSLILVNGANG	99	
QY	318	SDGSIIAILLIALLALWFWPLCC-----TVIIKEVPPP-----PAESEE	364	
Db	100	NDNTFIPL-----IHALTGGDTLCIARSADDTRIYRRIDPSENGNLVPVKQFRY	149	
QY	365	EDO---DGL-----PKKKWTVDASYGGRGVGIGIKMEVRW---GEKGSTEEG	407	
Db	150	TADLIYITGLDCLSGTISYSSDPTKKLYAISAG-----ITWSYGOGAVTWVG	195	
QY	408	AKL--EKAKNA-----RVKM-----PEQEYEFPEPRNLNNMMRRPS	441	
Db	196	ADIVADSKNTVALVTAAVVVQVTTSPPPPVRVSTSPPPPARVSSSSPPPATRSPPPRIT	255	
QY	442	SPRKWYSPIKGKLDAWLLVLLKRGYDRVSMVRPQPGDGTGRCTNFTRVKNKNQPAKYPLNNAY	501	
Db	256	SP----SPVLTAASPLPKTSPPPPRPVPPSPPPP-----VASPPPPPPPP--RV	297	



A:Accession: S28811  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-24,1189-1257,'S',1259-1263,'E',1265-1280 <TR>  
A:Cross-references: EMBL:X67327  
C:Genetics:  
C:Introns: 2845/3; 2863/3; 2887/3; 2899/3; 2922/1; 2985/1; 3008/1; 3065/1  
C:Superfamily: collagen alpha 1(XII) chain; fibronectin type III repeat homology; von Willebrand factor type A repeat homology; cell attachment (R-G-D) motif  
C:Keywords: alternative splicing; cell binding; coiled coil; connective tissue; disulfide bond; fibronectin type III repeat homology; collagen alpha 1(XII) chain short splice form #status predicted <MAN>  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-3124/Domain: collagen alpha 1(XII) chain #status predicted <MAN>  
F:24,1189-3124/Product: collagen alpha 1(XII) chain short splice form #status predicted  
F:24-1114/Domain: IIIA #status predicted <IIIA>  
F:24-105/Domain: fibronectin type III repeat homology <FN3A>  
F:137-301/Domain: von Willebrand factor type A repeat homology <VWAL>  
F:332-425/Domain: IIIB #status predicted <IIIB>  
F:332-414/Domain: fibronectin type III repeat homology <FN3B>  
F:629-601/Domain: von Willebrand factor type A repeat homology <VWA2>  
F:629-1178/Domain: IIIC #status predicted <IIIC>  
F:630-711/Domain: fibronectin type III repeat homology <FN3C>  
F:721-802/Domain: fibronectin type III repeat homology <FN3D>  
F:842-895/Domain: fibronectin type III repeat homology <FN3E>  
F:905-986/Domain: fibronectin type III repeat homology <FN3G>  
F:995-1076/Domain: fibronectin type III repeat homology <FN3H>  
F:1086-1169/Domain: fibronectin type III repeat homology <FN3I>  
F:1197-1361/Domain: von Willebrand factor type A repeat homology <VWA3>  
F:1384-2295/Domain: IIID #status predicted <IIID>  
F:1384-1465/Domain: fibronectin type III repeat homology <FN3J>  
F:1474-1557/Domain: fibronectin type III repeat homology <FN3K>  
F:1566-1647/Domain: fibronectin type III repeat homology <FN3L>  
F:1655-1738/Domain: fibronectin type III repeat homology <FN3M>  
F:1756-1838/Domain: fibronectin type III repeat homology <FN3N>  
F:1847-1928/Domain: fibronectin type III repeat homology <FN3O>  
F:1937-2019/Domain: fibronectin type III repeat homology <FN3P>  
F:2028-2110/Domain: fibronectin type III repeat homology <FN3Q>  
F:2119-2199/Domain: fibronectin type III repeat homology <FN3R>  
F:2207-2294/Domain: fibronectin type III repeat homology <FN3S>  
F:2325-2490/Domain: von Willebrand factor type A repeat homology <VWA4>  
F:2438-2440/Domain: cell adhesion #status predicted  
F:2438-2440/Region: cell adhesion #status predicted  
F:2509-2750/Domain: IXP, homologous to NC4 domain of type IX collagen #status predicted  
F:2751-2902/Domain: collagenous COL2 #status predicted <COL2>  
F:2899-2901/Region: cell attachment (R-G-D) motif  
F:2903-2945/Domain: non-collagenous NC2 #status predicted <NC2>  
F:2946-3048/Domain: collagenous COL1 #status predicted <COL1>  
F:3049-3324/Domain: non-collagenous NC1 #status predicted <NC1>  
F:32,1006,1032,1044,1512,1767,2210,2273,2532,2683/Binding site: carbohydrate (Asn) (covar)  
F:2780, 2789, 2836, 2842, 2860, 2866, 2869, 3004, 3007/Modified site: hydroxyproline (Pro) #stat

Query Match	6.6%	Score	200.5;	DB	1;	Length	3124;
Best Local Similarity	21.8%;	Pred.	No. 0.00012;				
Matches	144;	Conservative	76;	Mismatches	204;	Indels	237; Gaps
							34;
QY	39	CYGG-PDLYILDKSGV-LHHWNIIFYVEQLAHKF--ISPQ-LRMSEIFVFSTRGTTLM	93				
		: : : :         : :     :					
Db	2321	CRGAKAADIIVFLTDASWSIGDDNFNKVKVFVTNGAFDLINPAGIQVLSLVQVSDDAQSEF	2380				
		: : : :         : :     :					
QY	94	KLT--EDRQIRGLGLEELQKVLPGGDYMHGEFERASEQIY-YENRGYRTATSIALTLD	150				
		: :         : :     : :     :					
Db	2381	KLMTFDKAAQALGALNQV--YRGNTRTGKALTFFKEKVLTWESGRMGVPKVLVVVTD	2438				
		: :         : :     : :     :					
QY	151	GELHEDLFFYSEREANRSRDLAGIYVCVGKDFNFTQLARIAD--SKDHVFPYNDGFQAL	208				
		: :     : :     : :     : :     : :     : :     :					
Db	2439	GRSQDEV---RKAATVIQHSGFSVFVWGADVYNELAKIASPKSERHFIVDD-DFAD	2493				
		: :     : :     : :     : :     : :     : :     :					
QY	209	QGIIHTLKKSCTEIIAEEPSTICAGESQQVVVRGNFRHARNVDRLCVLCSFKINDSVTLN	268				
		: :     : :     : :     : :     : :     : :     :					
Db	2494	EKIQDNLVTFVTCETATSTCPLIYLEGYTSP-----GFMLESYNLT	2534				
		: :     : :     : :     : :     : :     : :     :					
QY	269	EKFPSVEDTYLLCPAPILKEVGVMKAALQ-VSMNDGL-----SFISSSVIIITTT	315				
		: :         : :     : :     : :     :					
Db	2535	EKHFE-----ASVOGVSLSEGSPFSVAVYRLHKNAFVSQP--IREI	2572				

Query Match 6.5%; Score 197.5; DB 2; Length 141;  
Best Local Similarity 58.3%; Pred. No. 3.2e-06;  
Matches 35; Conservative 2; Mismatches 18; Indels 5

[illegible]

RESULT 13

C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-Jun-1999  
C/Accession: T05225

C;Accession: T05225  
R;Bevan, M.; Vitale, D.; Liguori, R.; Argiriou, A.; De Sisti,  
submitted to the Protein Sequence Database July 1988

A; Accessio

A;Molecule type: DNA

A:Cross-references:

A: Experimental source: cultivar Columbia; BAC clone F17I5

C;Genetics:

A;Map position: 4

A;Note: F17I5.160

Query Match

Best Local Similar

Matches 47; Con:

.....

QY 473 PQPGDTGRCINFTRVKNNQPAKYPLNNAYHTSSPPPAPIYTPPPAP-----HCPP 5233
































Qv 524 pp----psApTppTpSpPSTL--pppp--QApPPNRAppP---SRPPPPRP 562

[illegible]

Db 552 pppvYSprrppppVHSprrppVFSprrppvYSprrppVHSprrppVHSprrppAP 601

## RESULT 14

C84534

hypothetical protein At.

C;Species: Arabidopsis thaliana (mouse-ear cress)

C:\Date: 02-Feb-2001  
C:\Accession: C091E34

C; Accession: C84334  
P: Lin Y : Kaul S :

K; LII, A.; Kaul, S.; Koo, H.; Moffat, K.

M.; ROO, H.; MULLAL, K.S.; CLOMIN, L.A.; SILEN, M.; VANAKEN, S.E.; OMAIYAN, M.; ROO, H.; MULLAL, K.S.; CLOMIN, L.A.; SILEN, M.; VANAKEN, S.E.; OMAIYAN, M.

Leuss, D.: J. NIELMAU,  
Nature 402 761-768

A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
nature 402, 701-706, 1999

A; Release  
A: Accession

A: Stat

A: Molecule type: DNA

A:Residues: 1-727 &lt;S

A:Cross-references: GB:AE002093; NID:q5306245; PIDN:AAD41978.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g15880

A;Map position: 2

1

Fri Aug 9 10:57:05 2002

Query Match	6.43;	Score 194;	DB 2;	Length 727;
Best Local Similarity	33.3%;	Pred. No. 4.6e-05;		
Matches	59;	Conservative	48;	Indels 54; Gaps 12;
		Mismatches	16;	

418	VKMQEYEQYEPEPEPLNNN	-----MRRPSSPKWYSP	IKGKLDALVLLRKGYDRVSWR	472
QY				
406	VHKQPPKESQPNFYNQSPVKFRSSPPQOQHVV	-----	VHS	446
DB				
473	PQPGD	---TGRCINFY--RVKNQZAK--	YPLNNAYHTS	518
QY				
447	PPASSPPTSPVHSTFSPVHKPQPKESQP	NDPYDQSPVKFRSSPPPVHSPSSPPS	505	
DB				
519	P-HCSPPPP	---SAPTPPTSP	-----STLPPPP-QAPPNRRAPP	560
QY				
				562

RESULT 15

SI14974  
 extensin class I (clone uG-18) - tomato (fragment)  
 C:Species: Lycopersicon esculentum (tomato)  
 C:Date: 07-May-1998 #sequence\_revision 15-May-1998 #text\_change 21-Jul-2000  
 C:Accession: SI4974  
 R:Shawalter, A.M.; Zhou, J.; Rumeau, D.; Worst, S.G.; Varner, J.E.  
 Plant Mol. Biol. 16, 547-565, 1991  
 A:Title: Tomato extensin and extensin-like cDNAs: structure and expression in response to wounding  
 A:Reference number: SI4970; MUID:91329690  
 A:Accession: SI4974  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-181 <SHO>  
 A:Cross-references: EMBL:X55685; NID:g19229; PIDN:CAA39215.1; PID:g1345537  
 A:Experimental source: cv. UC82B  
 C:Superfamily: hydroxyproline-rich glycoprotein  
 C:Keywords: cell wall; glycoprotein; hydroxyproline

Query Match	6.3%	Score 191.5;	DB 2;	Length 181;
Best Local Similarity	29.8%	Pred. No. 1.1e-05;		
Matches 51; Conservative	12;	Mismatches 35;	Indels 73;	Gaps 8;

421 PEQEYEFEPNLNNMRPSSPKWYSPKGLDALWLLRKGDRVSMRPQPGDTGR 480

Db 5 PPVYKSPPPPS-----PSPPFFII

QY: 481 CINFTRVKNNOQAKYPLNNAYHTSSPPAP 79

```
Db 31 -----PSPSPPPYYKSPPPPSPPSPFFFIINSFFFFSPPPIIINNNNNN-----
```

```

QY 528 APTPP-----IPSPSTLPPPP---QAPPPNRRAPPP-----SRPPFRFS 303
QY 528 APTPP-----IPSPSTLPPPP---QAPPPNRRAPPP-----SRPPFRFS 303

```

D6  
80 SPSPPPPYYKSPPPSPSPPPPPYYSSPPPPKKSPPPPPYYKSPPPPPSPS  
130

Search completed: August 9, 2002, 10:34:18  
Job time: 257 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: August 9, 2002, 10:46:21 ; Search time 102.68 Seconds  
(without alignments)  
950.225 Million cell updates/sec

Title: us-09-970-076-6  
Perfect score: 3025  
Sequence: 1 MATERRALGIGFOWLSLAT.....QAPPENRAPPSPRRPSPV 564

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1894	62.6	368	4 Q96P02	Q96P02 homo sapien
2	746.5	24.7	221	11 Q99L17	Q99L17 mus musculu
3	574	19.0	245	4 Q96NC7	Q96NC7 homo sapien
4	391.5	12.9	97	4 Q96EC6	Q96EC6 homo sapien
5	227	7.5	847	10 Q9XIB6	Q9XIB6 arabidopsis
6	213.5	7.1	687	10 Q948Y7	Q948Y7 volvox cart
7	213	7.0	956	10 Q9LJ64	Q9LJ64 arabidopsis
8	213	7.0	1143	10 Q948I6	Q948I6 volvox cart
9	211	7.0	409	10 Q9SBM1	Q9SBM1 volvox cart
10	211	7.0	1289	10 Q9FTQ7	Q9FTQ7 arabidopsis
11	205.5	6.8	760	10 Q9TOK5	Q9TOK5 arabidopsis
12	205	6.8	494	10 Q9FXA1	Q9FXA1 arabidopsis
13	204	6.7	919	6 Q28659	Q28659 arctolaqus
14	203.5	6.7	1151	5 Q9VAI4	Q9VAI4 oryctolaphila
15	203	6.7	599	10 P93797	P93797 volvox cart
16	202	6.7	464	10 Q41645	Q41645 volvox cart

17	201	6.6	579	10 Q9LGG8	Q9LGG8 oryza sativ
18	201	6.6	1201	10 Q9C6S1	Q9C6S1 arabidopsis
19	199.5	6.6	1392	10 Q9XER9	Q9XER9 arabidopsis
20	199	6.6	470	10 Q9LJL1	Q9LJL1 arabidopsis
21	198	6.5	242	10 Q43687	Q43687 vigna ungu
22	197.5	6.5	1059	10 Q94E61	Q94E61 oryza sativ
23	196.5	6.5	511	6 Q95JD0	Q95JD0 sus scrofa
24	196.5	6.5	566	6 Q95JD1	Q95JD1 sus scrofa
25	195.5	6.5	699	10 Q81765	Q81765 arabidopsis
26	194	6.4	727	10 Q9XIL9	Q9XIL9 arabidopsis
27	193.5	6.4	333	10 Q9XIZ3	Q9XIZ3 oryza sativ
28	193.5	6.4	787	3 Q94096	Q94096 pneumocysti
29	193	6.4	795	5 Q95UQ2	Q95UQ2 toxoplasma
30	193	6.4	938	3 Q96VI4	Q96VI4 pneumocysti
31	191.5	6.3	181	10 Q01943	Q01943 lycopersico
32	191.5	6.3	489	10 Q41707	Q41707 vigna ungu
33	191.5	6.3	676	6 Q95JC9	Q95JC9 sus scrofa
34	191.5	6.3	708	10 Q9SX31	Q9SX31 arabidopsis
35	191	6.3	322	10 Q09084	Q09084 lycopersico
36	190.5	6.3	134	10 Q9LT36	Q9LT36 arabidopsis
37	189.5	6.3	388	10 Q09082	Q09082 lycopersico
38	189	6.2	368	10 Q09085	Q09085 phaseolus v
39	188.5	6.2	158	12 Q9DWM0	Q9DWM0 pluteella xy
40	187.5	6.2	555	10 Q9FPQ6	Q9FPQ6 chlamydomon
41	186.5	6.2	1315	10 Q9SPM0	Q9SPM0 zea mays (m
42	186	6.1	192	10 Q65760	Q65760 cicer ariet
43	186	6.1	214	10 Q39789	Q39789 gossypium h
44	186	6.1	214	10 Q39763	Q39763 gossypium b
45	186	6.1	280	10 Q43682	Q43682 vigna ungu

## ALIGNMENTS

### RESULT 1

Q96P02 ID Q96P02 PRELIMINARY; PRT; 368 AA.  
AC Q96P02;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE ANTHRAX TOXIN RECEPTOR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bradley K.A., Mogridge J., Mourez M., Collier R.J., Young J.A.T.;  
RT "Identification Of the Cellular Receptor for Anthrax Toxin.";  
RL Nature 414:0-0(2001).  
DR EMBL; AF421380; AAL26496.1; -.  
KW Receptor.  
SQ SEQUENCE 368 AA; 41157 MW; 8A87B13FFA7D8753 CRC64;

Query Match	62.6%	Score	1894;	DB	4;	Length	368;
Best Local Similarity	100.0%	Pred. No.	2.4e-150;	Mismatches	0;	Indels	0;
Matches	364;	Conservative	0;	0;	0;	Gaps	0;
Qy	1	MATERRALGIGFOWLSLATLVLCAGQGGRRDGGPACYGFGFDLYFLDKSGSVLHHWN	60				
Db	1	MATERRALGIGFOWLSLATLVLCAGQGGRRDGGPACYGFGFDLYFLDKSGSVLHHWN	60				
Qy	61	EIYFVEQLAHKFISPOLRMSEFVSTGTTLMKLTEDREQIRQGLEELQKVLPGDITYM	120				
Db	61	EIYFVEQLAHKFISPOLRMSEFVSTGTTLMKLTEDREQIRQGLEELQKVLPGDITYM	120				
Qy	121	HEGFERASEQIYYENRQGYRTASVIAITDGLHEDLFFYSEREANRSDLGAIYVCV	180				
Db	121	HEGFERASEQIYYENRQGYRTASVIAITDGLHEDLFFYSEREANRSDLGAIYVCV	180				
Qy	181	KDFNETQLARIADSKDHVFPVNDGFGALQGIHLSILKKSCIEILAAEPSTICAGESFQV	240				

Fri Aug 9 10:57:06 2002

Db 181 KDNFTQLARIADSKDHVPVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQW 240  
 QY 241 VRNGFRHARNVDRLVCSFKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQVSNM 300  
 Db 241 VRNGFRHARNVDRLVCSFKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQVSNM 300  
 QY 301 DGLSFSTSSVITTTTSCDGSILAIALLILFLLALALWFWPCLCTVIIEVPPPPAE 360  
 Db 301 DGLSFSTSSVITTTTSCDGSILAIALLILFLLALALWFWPCLCTVIIEVPPPPAE 360  
 QY 361 ESEE 364  
 Db 361 ESEE 364

RESULT 2  
 Q99L17 PRELIMINARY; PRT; 221 AA.  
 AC Q99L17;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE HYPOTHETICAL 24.5 KDA PROTEIN (FRAGMENT).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RC TISSUE=MAMMARY TUMOR. WAP-TGF ALPHA MODEL. 7 MONTHS OLD, CROSS  
 RA Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC003908; AAH03908.1; -  
 KW Hypothetical protein.  
 FT NON\_TER  
 SQ SEQUENCE 221 AA; 24505 MW; 31AB702E42E9CEB3 CRC64;

Query Match 24.7%; Score 746.5; DB 11; Length 221;  
 Best Local Similarity 59.9%; Pred. No. 1.6e-54;  
 Matches 136; Conservative 30; Mismatches 44; Indels 17; Gaps 2;  
 QY 269 EXPFVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFSSVITTTTSCDGSILAIALL 328  
 Db 1 ERPVSIQPSILCPAPILVKNKGETLEVSISYNDGKSAYSRSLTITATECTINGIAIVAIL 60  
 QY 329 ILELLALALLWFWPCLCTVIIEVPPPPAESEEDDDGLPKKWTPTVDASYGGRGV 388  
 Db 61 VLLLLGALLWFWPCLCKVVIKDPPIPPPSAPMEEDDPLNKKWTPTVDASYGGRGV 120  
 QY 389 GGKRVKRGKSTEGEAGKLEKAKNARVKPEQYEFPEPRNLNNMRPSSPR--- 444  
 Db 121 GGKRVKRGKSTEGEAGKLEKAKNARVWVPEEIPIPS-----RPRPRPTHQ 171  
 QY 445 ----KWSPITGKLDALWLLRLKGYDRSVMRPQPGDTGRCINFTRV 487  
 Db 172 APQTKWTPIKGRDLALWALIMQYDRVSLMRPQEGDEGRINFTRV 218

RESULT 3  
 Q96NC7 PRELIMINARY; PRT; 245 AA.  
 AC Q96NC7;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE CDNA FLJ31074 FIS, CLONE HSYRA2001476.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;

SEQUENCE FROM N.A.  
 RN [1]  
 RP Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,  
 RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kanihara K.,  
 RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,  
 RA Otsuki T., Sato K., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,  
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
 RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;  
 RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; AK055636; BAB70976.1; -  
 SQ SEQUENCE 245 AA; 26111 MW; 61AE6EB0A2EBE506 CRC64;

Query Match 19.0%; Score 574; DB 4; Length 245;  
 Best Local Similarity 49.2%; Pred. No. 5.1e-40;  
 Matches 120; Conservative 45; Mismatches 67; Indels 12; Gaps 3;  
 QY 78 LRMSFVSTRTTLMKLTEDREQIRQGLELQKVLPGDVTYMHGEPERASEQIYVENRQ 137  
 Db 1 MRLSFVFSQATIIPLTGRGKISKGLDLKRVSPVGETIIEHGLKLANEQI--QKAG 58  
 QY 138 GYRTASVITATDGLHEDLFYISREANRSLGAIYVCYGVKDFNETQLARIADSKDH 197  
 Db 59 GLKTSIIIALTDGKLDGLVPSYAEKAKISRLGASVYCVGVLDDEQAQLERIADSKH 118  
 QY 198 VFPVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQVVRNGFRHARNVDRLV 257  
 Db 119 VFPVKGFOALQGIHSLKSCIEILAAEPSTICAGESFQVVRNGFRHARNVDRLV 178  
 QY 258 SPKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFSSV-ITTTTH 316  
 Db 179 TTYVNETYTSVKPVSVQLNSMLCPAPILNKAG-----EWGLTVTQAGVKWHDLTH 229  
 QY 317 CSDG 320  
 Db 230 CTFG 233

RESULT 4  
 Q96EC6 PRELIMINARY; PRT; 97 AA.  
 AC Q96EC6;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE UNKNOWN (PROTEIN FOR IMAGE:4705862) (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RC TISSUE=BREAST, AND MAMMARY ADENOCARCINOMA;  
 RA Strausberg R.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC012475; AAH12475.1; -  
 FT NON\_TER  
 SQ SEQUENCE 97 AA; 10453 MW; 14F475F0B170E71A CRC64;

Query Match 12.9%; Score 391.5; DB 4; Length 97;  
 Best Local Similarity 90.0%; Pred. No. 2.6e-25;  
 Matches 81; Conservative 1; Mismatches 5; Indels 3; Gaps 1;  
 QY 237 FOVYVRNGFRHARNVDRLVCSFKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQ 296  
 Db 1 FOVYVRNGFRHARNVDRLVCSFKINDSVTLNEKPSVEDTYLLCPAPILKEVGMKAALQ 60  
 QY 297 VSMNDGLSFSSVITTTTSCDGSILAI 326  
 Db 1 VSMNDGLSFSSVITTTTSCDGSILAI 326



Db 513 KOBPPKPEESKPPKPEESKPPKQETPKPEES-----PKQPPKQETPKP 562  
Qy 405 EEGAKLEKAKNARVKMEQYEFEPENLNNMRPSSPRKWSPTIKGDLALWLLRRG 464  
Db 563 EESPKQPPKQETPK-PE---ESPKEQ-----PPKQEQPKTEAPKMGSPPLESPVPNDP 613  
Qy 465 YDR--VSVMRPQGDGRCINFRVKNNOAKYPLNN-----AYHT 503  
Db 614 YDASPIKKRRQPPSP---STEEKTTSPQSPVHSPPPPPVHSPPPVFSPPPMHS 669  
Qy 504 SSP-----PPAPIYTPPPHPCPPPPSAPTPIPSPPSTL-----PPPP--QAPP 548  
Db 670 PPPPVYSPPVHSPPPPPVHSPPPVHSPPPVHSPPPVHSPPPVHSPPPVHSPPP 729  
Qy 549 PNRAPPD---SRPPRP 562  
Db 730 PVQSPPPPPVFSPPPPAP 747

Query Match 7.0%; Score 211; DB 10; Length 409;  
Best Local Similarity 51.9%; Pred. NO. 2.7e-09;  
Matches 40; Conservative 3; Mismatches 28; Indels 6; Gaps 2;

Qy 492 PAKYPLNNAYHTSSPPAPI-----YTPPPHPCPPPPSAPTPIPSPPSTLPPPPQA 546  
Db 76 PPQPLPPSPSPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 134  
Qy 547 PPNRAPPPSRPPRPS 563  
Db 135 PPNRAPPPSRPPRPS 151

RESULT 10

Q9FLQ7 PRELIMINARY; PRT; 1289 AA.  
ID Q9FLQ7  
AC Q9FLQ7  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE GB/AAD23008.1.  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX MEDLINE=98290546; PubMed=9628582;  
RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,  
Tabata S.;  
RA "Structural analysis of Arabidopsis thaliana chromosome 5. IV.  
RT Sequence features of the regions of 1,456,315 bp covered by nineteen  
RT physically assigned P1 and TAC clones.";  
RL DNA Res. 5:41-54(1998).  
DR EMBL: AB010070; BAB11454.1;  
DR InterPro: IPR002965; P-rich.extensn.  
DR PRINTS: PR01217; PRICHEXTENS.  
SQ SEQUENCE 1289 AA; 137558 MW; 898E55A2A618F5E0 CRC64;

Query Match 7.0%; Score 211; DB 10; Length 1289;  
Best Local Similarity 23.9%; Pred. No. 1.4e-08;  
Matches 74; Conservative 24; Mismatches 91; Indels 120; Gaps 12;

Qy 350 IIEKVPVPPAESEEDDGL-----PKKKWPTVDAS----- 381  
Db 708 VLPVPPPPPLPFSERPNSGTVLPVPPPPPPPKWSVYASALAIPTCSTSQAPTSPTPPPP 767  
Qy 382 ---YGGVGGVGIKRMVVRMEVKGSTEGAKLEK-----AKNARVKMPEOEY 425  
Db 768 PPATYS-----VQKSSDLQTSQLPSPPPPPPPPPPPPPPPPPPPPPPPPPPPP 814

Db 513 KOBPPKPEESKPPKPEESKPPKQETPKPEES-----PKQPPKQETPKP 562  
Qy 405 EEGAKLEKAKNARVKMEQYEFEPENLNNMRPSSPRKWSPTIKGDLALWLLRRG 464  
Db 563 EESPKQPPKQETPK-PE---ESPKEQ-----PPKQEQPKTEAPKMGSPPLESPVPNDP 613  
Qy 465 YDR--VSVMRPQGDGRCINFRVKNNOAKYPLNN-----AYHT 503  
Db 614 YDASPIKKRRQPPSP---STEEKTTSPQSPVHSPPPPPVHSPPPVFSPPPMHS 669  
Qy 504 SSP-----PPAPIYTPPPHPCPPPPSAPTPIPSPPSTL-----PPPP--QAPP 548  
Db 670 PPPPVYSPPVHSPPPPPVHSPPPVHSPPPVHSPPPVHSPPPVHSPPPVHSPPP 729  
Qy 549 PNRAPPD---SRPPRP 562  
Db 730 PVQSPPPPPVFSPPPPAP 747

Query Match 7.0%; Score 213; DB 10; Length 1143;  
Best Local Similarity 51.9%; Pred. No. 8.1e-09;  
Matches 40; Conservative 5; Mismatches 30; Indels 2; Gaps 1;

Qy 487 VKNNOAKYPLNNAYHTSSPPAPIYTPPPHPCPPPPSAPTPIPSPPSTLPPPPQA 546  
Db 492 VKTSPVFPVPPPPSPLTGRPPSP--RPPRSPSPPPPPPPPPPPPPPPPPPPPPSP 549  
Qy 547 PPNRAPPPSRPPRPS 563  
Db 550 PPNRAPPPSRPPRPS 566

RESULT 9

Q9SBM1 PRELIMINARY; PRT; 409 AA.  
ID Q9SBM1  
AC Q9SBM1  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE HYDROXYPROLINE-RICH GLYCOPROTEIN DZ-HRGP PRECURSOR.  
GN HRGP GENE.  
OS Volvox carteri f. nagariensis.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
OC Volvocaceae; Volvox.  
OX NCBI\_TaxID=3068;  
RN [1]





Db 398 HTIEPPPPAPPTLVPPPPAPPTIKPPPPAPPTVEPPPPPPPPAPPTVEPPPPPPAPPT 457  
Qy 551 RAPPSPRPPRPS 563  
Db 458 KVEPP--PPPAFA 468

RESULT 15  
P93797  
ID P93797 PRELIMINARY; PRT: 599 AA.  
AC P93797;  
DT 01-MAY-1997 (TREMBlrel. 03, Created)  
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE PHEROPHORIN-S, PRECURSOR.  
OS Volvox carteri.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
OC Volvocaceae; Volvox.  
OX NCBI\_TaxID=3067;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HK 10;  
RX MEDLINE=97162277; Pubmed=9009264;  
RA Godl K., Hallmann A., Wenzl S., Sumper M.;  
RT "Differential targeting of closely related ECM-glycoproteins: The  
RT pherophorin family from Volvox.";  
RL EMBO J. 16:25-34(1997).  
DR EMBL; Y07752; CAA69032.1; -.  
KW Signal.  
FT SIGNAL 1 17 POTENTIAL.  
FT CHAIN 18 599 POTENTIAL.  
SQ SEQUENCE 599 AA; 63436 MW; 8BAA14D5A71F8082 CRC64;

Query Match 6.7%; Score 203; DB 10; Length 599;  
Best Local Similarity 55.9%; Pred. No. 2.2e-08;  
Matches 38; Conservative 1; Mismatches 25; Indels 4; Gaps 1;  
Qy 496 PLNNAYHTSSPPAPITPPPPAPHCPPPPSAPTPPSPSTLPPPPQAPPPNRAPPP 555  
Db 213 PLPNA-----PPSPLPPSPPPPPPPSPPPPPPPPPPPPPPPPPPPPPPPPP 268  
Qy 556 SRPPRPS 563  
Db 269 PPPPPPS 276

Search completed: August 9, 2002, 10:46:25  
Job time: 848 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:33:05 ; Search time 45.48 Seconds  
(without alignments)  
302.903 Million cell updates/sec

Title: US-09-970-076-6  
Perfect score: 3025  
Sequence: 1 MATARRALGIGFQWLSLAT.....QAPPPNRAPPPRRPPSV 564

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PC10S\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	191.5	6.3	306	1	US-08-217-327-6
2	186	6.1	214	1	US-08-217-327-4
3	173.5	5.7	334	6	5202236-3
4	171.5	5.7	331	6	5202236-37
5	167.5	5.5	1248	2	US-09-080-897-2
6	167.5	5.5	1248	4	US-09-323-735-2
7	166	5.5	1155	1	US-08-286-889-46
8	166	5.5	1155	1	US-08-485-618-46
9	166	5.5	1155	1	US-08-362-652-46
10	166	5.5	1155	2	US-08-605-672-46
11	166	5.5	1155	2	US-08-482-293A-46
12	166	5.5	1155	2	US-08-943-363-46
13	166	5.5	1155	4	US-09-193-043-46
14	166	5.5	1161	1	US-08-485-618-53
15	166	5.5	1161	1	US-08-362-652-53
16	166	5.5	1161	2	US-08-605-672-53
17	166	5.5	1161	2	US-08-482-293A-53
18	166	5.5	1161	2	US-08-943-363-53
19	166	5.5	1161	4	US-09-193-043-53
20	163	5.4	635	1	US-07-832-855-2
21	161.5	5.3	1315	3	US-08-899-595-3
22	160.5	5.3	1255	2	US-09-080-897-4
23	160.5	5.3	1255	3	US-08-899-595-1
24	160.5	5.3	1255	4	US-09-323-735-4
25	159.5	5.3	1151	1	US-08-286-889-37
26	159.5	5.3	1151	1	US-08-485-618-37
27	159.5	5.3	1151	1	US-08-362-652-37

28	159.5	5.3	1151	2	US-08-605-672-37	Sequence 37, Appl
29	159.5	5.3	1151	2	US-08-482-293A-37	Sequence 37, Appl
30	159.5	5.3	1151	2	US-08-943-363-37	Sequence 37, Appl
31	159.5	5.3	1151	4	US-09-193-043-37	Sequence 37, Appl
32	159.5	5.3	1161	1	US-08-173-497-2	Sequence 2, Appl
33	159.5	5.3	1161	1	US-08-286-889-2	Sequence 2, Appl
34	159.5	5.3	1161	1	US-08-485-618-2	Sequence 2, Appl
35	159.5	5.3	1161	1	US-08-485-618-55	Sequence 55, Appl
36	159.5	5.3	1161	1	US-08-485-618-99	Sequence 99, Appl
37	159.5	5.3	1161	1	US-08-362-652-2	Sequence 2, Appl
38	159.5	5.3	1161	1	US-08-362-652-55	Sequence 55, Appl
39	159.5	5.3	1161	2	US-08-605-672-2	Sequence 2, Appl
40	159.5	5.3	1161	2	US-08-605-672-55	Sequence 55, Appl
41	159.5	5.3	1161	2	US-08-605-672-99	Sequence 99, Appl
42	159.5	5.3	1161	2	US-08-482-293A-2	Sequence 2, Appl
43	159.5	5.3	1161	2	US-08-482-293A-55	Sequence 55, Appl
44	159.5	5.3	1161	2	US-08-482-293A-99	Sequence 99, Appl
45	159.5	5.3	1161	2	US-08-943-363-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-217-327-6  
; Sequence 6, Application US/08217327  
; Patent No. 5474925  
; GENERAL INFORMATION:  
; APPLICANT: John, Maliyakal E  
; APPLICANT: Barton, Kenneth A  
; TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles and Brady  
; STREET: P.O. Box 2113  
; CITY: Madison  
; STATE: WI  
; COUNTRY: USA  
; ZIP: 53701-2113  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/217,327  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/812,233  
; FILING DATE: 19-DEC-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27,386  
; REFERENCE/DOCKET NUMBER: 1122990831  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 306 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-217-327-6

Query Match 6.3%; Score 191.5; DB 1; Length 306;  
Best Local Similarity 26.1%; Pred. No. 4.8e-09;  
Matches 59; Conservative 18; Mismatches 56; Indels 93; Gaps 10;  
QY 352 KEVPPPP---PAESEEEDDDGLPKKKVPTDASYYGGRGVGKIKRMEVRWKEGKSTEEGA 408  
| ||| ||| : | | : | | : |

Fri Aug 9 10:57:04 2002

Db 144 KSPPPKHPAHHYKYSPPPKHPPAHHY-----177  
QY 409 KLEKAKNARVMQEEVEFPEPRNLNNMRSPRKWYSPKIGKLDALWLLRKGYDRV 468  
Db 178 -KYKYSPPPTPVYKYSPP-----PTPVYKYSKSP-----208  
QY 469 SVMRQPGDTCRCINFTRVKNNOBAPKYPPLANNAYHTSSPPPAPIYTPPPAPHCPPPPPP--526  
Db 209 ----PPP-----KHSPA--PVHHYKYSPPPTPVYKYSPPPEHSPPTPV 249  
QY 527 ---SAPTPPSPSTLPPPP-----QAPPPNRAPP--SRPPR 561  
Db 250 YKYSPPPPMHSP--PPTPVYKYSPPPPMHSPPPPVYSPPPK 292  
RESULT 2  
US-08-217-327-4  
; Sequence 4, Application US/08217327  
; Patent No. 5474925  
; GENERAL INFORMATION:  
; APPLICANT: John, Maliyakal E  
; APPLICANT: Barton, Kenneth A  
; TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles and Brady  
; STREET: P.O. Box 2113  
; CITY: Madison  
; STATE: WI  
; COUNTRY: USA  
; ZIP: 53701-2113  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/217,327  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/812,233  
; FILING DATE: 19-DEC-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27,386  
; REFERENCE/DOCKET NUMBER: 1122990831  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 214 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-217-327-4  
Query Match 6.1%; Score 186; DB 1; Length 214;  
Best Local Similarity 47.0%; Pred. No. 8.7e-09;  
Matches 39; Conservative 5; Mismatches 23; Indels 16; Gaps 4;  
QY 496 PLNNAYHTSSPPA-----PIYTPPP--PAPHCPPP--PPSAPTPPSPSTLPP-----542  
Db 60 PVSTPPPTSSPPPVASPPVSTTPSSPPPTATPPASPPPTATPPASPPPTATPPASPP 119  
QY 543 ---PPQAPPNNRAPPSPRPPR 562  
Db 120 PATPPPTATPPATPPATPP 142  
RESULT 3

5202236-3  
; Patent No. 5202236  
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,  
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILPULA, DAVID  
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE  
; PROTEIN  
; NUMBER OF SEQUENCES: 39  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/528,762  
; FILING DATE: 23-MAY-1990  
; APPLICATION NUMBER: 82,456  
; FILING DATE: 07-AUG-1987  
; APPLICATION NUMBER: 933,945  
; FILING DATE: 24-NOV-1986  
; APPLICATION NUMBER: 650,128  
; FILING DATE: 13-SEP-1984  
; SEQ ID NO: 3:  
; LENGTH: 334  
5202236-3  
Query Match 5.7%; Score 173.5; DB 6; Length 334;  
Best Local Similarity 51.6%; Pred. No. 2.1e-07;  
Matches 33; Conservative 4; Mismatches 16; Indels 11; Gaps 3;  
QY 507 PPAPITYTPPP-----PAPHCPPPPPSAPTPPPIPSPTLPPPPQAPPNNRAPPSPR 558  
Db 8 PPAPAPAPALAPAPNPNPNQSPSPPTTP--PTPPS--PPAPSPSPSPNPPSP 64  
QY 559 PPRP 562  
Db 65 PPSP 68  
RESULT 4  
5202236-37  
; Patent No. 5202236  
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,  
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILPULA, DAVID  
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE  
; PROTEIN  
; NUMBER OF SEQUENCES: 39  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/528,762  
; FILING DATE: 23-MAY-1990  
; APPLICATION NUMBER: 82,456  
; FILING DATE: 07-AUG-1987  
; APPLICATION NUMBER: 933,945  
; FILING DATE: 24-NOV-1986  
; APPLICATION NUMBER: 650,128  
; FILING DATE: 13-SEP-1984  
; SEQ ID NO: 37:  
; LENGTH: 331  
5202236-37  
Query Match 5.7%; Score 171.5; DB 6; Length 331;  
Best Local Similarity 51.6%; Pred. No. 3.1e-07;  
Matches 33; Conservative 4; Mismatches 16; Indels 11; Gaps 3;  
QY 507 PPAPITYTPPP-----PAPHCPPPPPSAPTPPPIPSPTLPPPPQAPPNNRAPPSPR 558  
Db 5 PPAPAPAPALAPAPNPNPNQSPSPPTTP--PTPPS--PPAPSPSPSPNPPSP 61  
QY 559 PPRP 562  
Db 62 PPSP 65  
RESULT 5  
US-09-080-897-2  
; Sequence 2, Application US/09080897  
; Patent No. 5985574



us-09-970-076-6.ra1

Fri Aug 9 10:57:04 2002

```

; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32797
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1155 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-485-618-46

Query Match 5.5%; Score 166; DB 1; Length 1155;
Best Local Similarity 28.2%; Pred. No. 5.9e-06;
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 37 PACYG-GFDLYFILDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMK 94
Db 144 PECPGQEMDIAFLDGSIDQSDFTQMKDFVKALMGQLASTSTFSLMOYSNLIKTHFT 203
QY 95 LTEDR-----EQIRGLEELQKVLPGDVTYMHGFEFASQIYYENRQGYRTA-SVIAL 148
Db 204 FTEFKSSLSQSLDAIVQLQ-----GLTYTASGIQKVVKELFHSKNGARKSAKKILIVI 258
QY 149 TDGELHEDLFYSE--REANRSDLGAIYVCVVKD-FNE-TQLARI-----ADSKDHVF 199
Db 259 TDGQKFRDPLEYRHVPEAKA---GIIRYATGVGDFAFREPTALQELNTIGSAPSODHVF 315
QY 200 PVDGFOALQGIHLSILKKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVD 253
Db 316 KVGN-FVALRSIQRIQEK----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

RESULT 9
US-08-362-652-46
; Sequence 46, Application US/08362652
; Patent No. 5766850
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,652
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994

```

```

; FILING DATE: 435
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: P38,659
; REFERENCE/DOCKET NUMBER: 27866/32168
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1155 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-286-889-46

Query Match 5.5%; Score 166; DB 1; Length 1155;
Best Local Similarity 28.2%; Pred. No. 5.9e-06;
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 37 PACYG-GFDLYFILDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMK 94
Db 144 PECPGQEMDIAFLDGSIDQSDFTQMKDFVKALMGQLASTSTFSLMOYSNLIKTHFT 203
QY 95 LTEDR-----EQIRGLEELQKVLPGDVTYMHGFEFASQIYYENRQGYRTA-SVIAL 148
Db 204 FTEFKSSLSQSLDAIVQLQ-----GLTYTASGIQKVVKELFHSKNGARKSAKKILIVI 258
QY 149 TDGELHEDLFYSE--REANRSDLGAIYVCVVKD-FNE-TQLARI-----ADSKDHVF 199
Db 259 TDGQKFRDPLEYRHVPEAKA---GIIRYATGVGDFAFREPTALQELNTIGSAPSODHVF 315
QY 200 PVDGFOALQGIHLSILKKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVD 253
Db 316 KVGN-FVALRSIQRIQEK----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

RESULT 8
US-08-485-618-46
; Sequence 46, Application US/08485618
; Patent No. 5728533
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,618
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889

```

ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32391  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1155 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-362-652-46

Query Match 5.5%; Score 166; DB 1; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 5.9e-06;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;  
Qy 37 PACYG-GFDLYFLDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFIVFSTGRGTTLMK 94  
Db 144 PECQEQMDIAFLIDGSGIDQSDFTQMKDFVKALMGQLASTSTFSLMQYSNLIKTHFT 203  
Qy 95 LTEDR-----EQIRGLELEQKVLPGGDTYMHGFERASEQIYYENRQGYRTA-SVITIAL 148  
Db 204 FTEFKSSLSQSLVDIAVQLQ-----GLTYTASGIQKVVKELFHSKNGARKSAKKILIVI 258  
Qy 149 TDGELHEDLFFYSE--REANRSRDIGAIVYGVVKD-FNE-TQLARI-----ADSKDHVF 199  
Db 259 TDGQKFRDPLEYRHRVPEAKA--GIIRYAIGVGDAFREPTALQELNTIGSAPSQDHVF 315  
Qy 200 PVNDGFOALGIIHSILKSKCIEILAAEPSTICAGESFQVVVVRGNGFRHARNVD 253  
Db 316 KVGN-FVALRSIQRIQEK----IFAIEGTSRSSSFQHEMSQEGFSALSMD 364

RESULT 10  
US-08-605-672-46  
Sequence 46, Application US/08605672  
Patent No. 5817515  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/605,672  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:

NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1155 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-605-672-46

Query Match 5.5%; Score 166; DB 2; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 5.9e-06;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;  
Qy 37 PACYG-GFDLYFLDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFIVFSTGRGTTLMK 94  
Db 144 PECQEQMDIAFLIDGSGIDQSDFTQMKDFVKALMGQLASTSTFSLMQYSNLIKTHFT 203  
Qy 95 LTEDR-----EQIRGLELEQKVLPGGDTYMHGFERASEQIYYENRQGYRTA-SVITIAL 148  
Db 204 FTEFKSSLSQSLVDIAVQLQ-----GLTYTASGIQKVVKELFHSKNGARKSAKKILIVI 258  
Qy 149 TDGELHEDLFFYSE--REANRSRDIGAIVYGVVKD-FNE-TQLARI-----ADSKDHVF 199  
Db 259 TDGQKFRDPLEYRHRVPEAKA--GIIRYAIGVGDAFREPTALQELNTIGSAPSQDHVF 315  
Qy 200 PVNDGFOALGIIHSILKSKCIEILAAEPSTICAGESFQVVVVRGNGFRHARNVD 253  
Db 316 KVGN-FVALRSIQRIQEK----IFAIEGTSRSSSFQHEMSQEGFSALSMD 364

RESULT 11  
US-08-482-293A-46  
Sequence 46, Application US/08482293A  
Patent No. 5831029  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,293A  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.

us-09-970-076-6.rai

Fri Aug 9 10:57:04 2002

REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1155 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-943-363-46

Query Match 5.5%; Score 166; DB 2; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 5.9e-06;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;  
QY 37 PACYG-GFDLYFILDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFVIFSTRGTTLMK 94  
Db 144 PECPCQEMDIAFLIDGSGSIDQSDFTQMKDFVKALMGQLASTSTFSLMQYSNLIKTHFT 203  
QY 95 LTEDR-----EQIROGLELEQKVLPGGDTYMHGFEFRASEQIYYENRQGYRTA-SVITIAL 148  
Db 204 FTEFKSSLSPOSILVDAIVOLQ-----GLTYTAGIQQVVKELFHSKNGARKSAKKILIVI 258  
QY 149 TDGELHEDLFFYSE--REANRSRDLGAIYVCVGVKD-FNE-TQLARI-----ADSKDHVF 199  
Db 259 TDGQKFRDPLEYRHVPIAEKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSODHVF 315  
QY 200 PVNDGQALQGIHLSILKKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVD 253  
Db 316 KVGNG-FVALRSIQRIQEK-----IFAIEGTESRSSSFQHEMSQEGFSALSMD 364

## RESULT 13

US-09-193-043-46  
Sequence 46, Application US/09193043  
Patent No. 6251395  
GENERAL INFORMATION:  
APPLICANT: Gallatin, Michael W.  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 6251395el Human 2  
FILE REFERENCE: 27866/35004  
CURRENT APPLICATION NUMBER: US/09/193,043  
CURRENT FILING DATE: 1998-11-16  
EARLIER APPLICATION NUMBER: 08/173,497  
EARLIER FILING DATE: 1993-12-23  
EARLIER APPLICATION NUMBER: 08/286,889  
EARLIER FILING DATE: 1994-08-05  
EARLIER APPLICATION NUMBER: 08/362,652  
EARLIER FILING DATE: 1994-12-21  
EARLIER APPLICATION NUMBER: 08/943,363  
EARLIER FILING DATE: 1997-10-03  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 46  
LENGTH: 1155  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-193-043-46

Query Match 5.5%; Score 166; DB 4; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 5.9e-06;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 37 PACYG-GFDLYFILDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFVIFSTRGTTLMK 94  
Db 144 PECPCQEMDIAFLIDGSGSIDQSDFTQMKDFVKALMGQLASTSTFSLMQYSNLIKTHFT 203  
QY 95 LTEDR-----EQIROGLELEQKVLPGGDTYMHGFEFRASEQIYYENRQGYRTA-SVITIAL 148

REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1155 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-482-293A-46

Query Match 5.5%; Score 166; DB 2; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 5.9e-06;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;  
QY 37 PACYG-GFDLYFILDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFVIFSTRGTTLMK 94  
Db 144 PECPCQEMDIAFLIDGSGSIDQSDFTQMKDFVKALMGQLASTSTFSLMQYSNLIKTHFT 203  
QY 95 LTEDR-----EQIROGLELEQKVLPGGDTYMHGFEFRASEQIYYENRQGYRTA-SVITIAL 148  
Db 204 FTEFKSSLSPOSILVDAIVOLQ-----GLTYTAGIQQVVKELFHSKNGARKSAKKILIVI 258  
QY 149 TDGELHEDLFFYSE--REANRSRDLGAIYVCVGVKD-FNE-TQLARI-----ADSKDHVF 199  
Db 259 TDGQKFRDPLEYRHVPIAEKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSODHVF 315  
QY 200 PVNDGQALQGIHLSILKKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVD 253  
Db 316 KVGNG-FVALRSIQRIQEK-----IFAIEGTESRSSSFQHEMSQEGFSALSMD 364

## RESULT 12

US-08-943-363-46  
Sequence 46, Application US/08943363  
Patent No. 5837478  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 114  
CORRESPONDENCE ADDRESS: 114  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/943,363  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659

Db 204 FTEFKSSLSQSLVDAIVQLQ-----GLTYTASGIQKVVKELFHSKNGARKSAKKILIVI 258  
QY 149 TDGELHEDLFFYSE--REANRSRDGAIVYCVGVKD-FNE-TOLARI-----ADSKDHVF 199  
Db 259 TDGQKFRDPLEYRHVPEAEKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSQDHVF 315  
QY 200 PVNDGFALQGIHSLKSCIEILAAEPSTICAGESFQVVYRNGRFRHARNVD 253  
Db 316 KVG-N-FVALRSIQRIQIEK----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

RESULT 14  
US-08-485-618-53  
; Sequence 53, Application US/08485618  
; Patent No. 5728533  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485,618  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32797  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1161 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-485-618-53

Query Match 5.5%; Score 166; DB 1; Length 1161;  
Best Local Similarity 28.2%; Pred. No. 5.9e-06;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;  
QY 37 PACYG-GFDLYFILDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMK 94  
Db 144 PECPCQEMDIAFLIDGSSIDSDFTQMKDFVKALMGQLASTSTFSFLMQYSNLKTHFT 203  
QY 95 LTEDR-----EQIROGLEELQKVLPGGDTYMHGEPERASEQIYYENRGYRTA-SVIAL 148  
Db 204 FTEFKSSLSQSLVDAIVQLQ-----GLTYTASGIQKVVKELFHSKNGARKSAKKILIVI 258

QY 149 TDGELHEDLFFYSE--REANRSRDGAIVYCVGVKD-FNE-TOLARI-----ADSKDHVF 199  
Db 259 TDGQKFRDPLEYRHVPEAEKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSQDHVF 315  
QY 200 PVNDGFALQGIHSLKSCIEILAAEPSTICAGESFQVVYRNGRFRHARNVD 253  
Db 316 KVG-N-FVALRSIQRIQIEK----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364  
RESULT 15  
US-08-362-652-53  
; Sequence 53, Application US/08362652  
; Patent No. 5766850  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/362,652  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32391  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1161 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-362-652-53

Query Match 5.5%; Score 166; DB 1; Length 1161;  
Best Local Similarity 28.2%; Pred. No. 5.9e-06;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;  
QY 37 PACYG-GFDLYFILDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMK 94  
Db 144 PECPCQEMDIAFLIDGSSIDSDFTQMKDFVKALMGQLASTSTFSFLMQYSNLKTHFT 203  
QY 95 LTEDR-----EQIROGLEELQKVLPGGDTYMHGEPERASEQIYYENRGYRTA-SVIAL 148  
Db 204 FTEFKSSLSQSLVDAIVQLQ-----GLTYTASGIQKVVKELFHSKNGARKSAKKILIVI 258  
QY 149 TDGELHEDLFFYSE--REANRSRDGAIVYCVGVKD-FNE-TOLARI-----ADSKDHVF 199  
Db 259 TDGQKFRDPLEYRHVPEAEKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSQDHVF 315





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:35:55 ; Search time 98.32 Seconds  
(without alignments)  
525.968 Million cell updates/sec

Title: us-09-970-076-2\_COPY\_27\_321

Perfect score: 1536

Sequence: 1 GGGRRDGGPACGGFDLY.....GLSFSSVITTHCSGDS 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 528882 seqs, 175299045 residues

Total number of hits satisfying chosen parameters: 528882

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_New.\*

- 1: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1536	100.0	564	1	PCT-US02-08253-187 Sequence 187, App
2	1536	100.0	564	1	PCT-US02-08253-232 Sequence 232, App
3	1536	100.0	564	7	US-60-389-987-1823 Sequence 1823, Ap
4	1532	99.7	551	6	US-10-038-307-18 Sequence 18, Appl
5	1520	99.0	333	6	US-10-038-307-2 Sequence 2, Appl
6	1520	99.0	345	6	US-10-038-307-20 Sequence 24, Appl
7	1520	99.0	564	6	US-10-038-307-20 Sequence 20, Appl
8	1519	98.9	562	1	PCT-US02-08253-194 Sequence 194, App
9	1519	98.9	562	1	PCT-US02-08253-301 Sequence 301, App
10	1507	98.1	328	6	US-10-038-307-26 Sequence 26, Appl
11	1505.5	98.0	342	6	US-10-038-307-22 Sequence 22, Appl
12	1500	97.7	543	6	US-10-038-307-14 Sequence 14, Appl
13	1500	97.7	543	6	US-10-038-307-16 Sequence 16, Appl
14	1490	97.0	543	6	US-10-038-307-10 Sequence 10, Appl
15	1479	96.3	534	6	US-10-038-307-12 Sequence 12, Appl
16	966	62.9	218	5	US-09-629-469A-11706 Sequence 11706, A
17	903	58.8	538	6	US-10-047-542-99 Sequence 99, Appl
18	790.5	51.5	488	6	US-10-104-047-2639 Sequence 2639, Ap
19	790.5	51.5	488	6	US-10-038-307-6 Sequence 6, Appl
20	790.5	51.5	488	7	US-60-373-595-10 Sequence 10, Appl
21	777.5	50.6	587	6	US-10-125-540-312 Sequence 312, App
22	770	50.1	487	6	US-10-038-307-8 Sequence 8, Appl
23	685	44.6	381	6	US-10-038-307-4 Sequence 4, Appl
24	145.5	9.5	1152	5	US-09-592-617A-43 Sequence 43, Appl
25	138.5	9.0	1163	1	PCT-US02-10824-116 Sequence 116, App
26	138.5	9.0	1163	7	US-60-369-452-33 Sequence 33, Appl

27	132.5	8.6	191	5	US-09-805-354-1	Sequence 1, Appl
28	131	8.5	3063	1	PCT-US02-08253-257	Sequence 257, App
29	131	8.5	3063	1	PCT-US02-19669-61	Sequence 61, Appl
30	131	8.5	3063	1	PCT-US02-19669-63	Sequence 63, Appl
31	131	8.5	3063	6	US-10-177-293-61	Sequence 61, Appl
32	131	8.5	3063	6	US-10-177-293-63	Sequence 63, Appl
33	128.5	8.4	646	6	US-10-155-881-22380	Sequence 22380, A
34	128	8.3	3051	7	US-60-360-039-5866	Sequence 5866, Ap
35	126	8.2	460	6	US-10-155-881-33525	Sequence 33525, A
36	125.5	8.2	191	5	US-09-805-354-2	Sequence 2, Appl
37	124.5	8.1	191	5	US-09-805-354-3	Sequence 3, Appl
38	121.5	7.9	187	5	US-09-592-617A-59	Sequence 59, Appl
39	118.5	7.7	3594	6	US-10-150-821-4	Sequence 4, Appl
40	117.5	7.6	1207	7	US-60-389-987-1591	Sequence 1591, Ap
41	114	7.4	682	6	US-10-155-881-28757	Sequence 28757, A
42	113.5	7.4	764	1	PCT-US02-19669-27	Sequence 27, Appl
43	113.5	7.4	764	6	US-10-177-293-27	Sequence 27, Appl
44	113.5	7.4	795	6	US-10-009-557-7	Sequence 7, Appl
45	113.5	7.4	915	5	US-09-904-920A-34	Sequence 34, Appl

## ALIGNMENTS

RESULT 1

PCT-US02-08253-187

; Sequence 187, Application PC/TUS0208253

; GENERAL INFORMATION:

; APPLICANT: Carson-Walter, Eleanor

; APPLICANT: St. Croix, Brad

; APPLICANT: Vogelstein, Bert

; APPLICANT: Kinzler, Kenneth

; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS

; FILE REFERENCE: 1107.00179

; CURRENT APPLICATION NUMBER: PCT/US02/08253

; CURRENT FILING DATE: 2002-04-10

; PRIOR APPLICATION NUMBER: 60/282,850

; PRIOR FILING DATE: 2001-04-11

; PRIOR APPLICATION NUMBER: 60/308,829

; PRIOR FILING DATE: 2001-08-01

; NUMBER OF SEQ ID NOS: 359

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 187

; LENGTH: 564

; TYPE: PRT

; ORGANISM: Homo sapiens

PCT-US02-08253-187

Query Match 100.0%; Score 1536; DB 1; Length 564;  
Best Local Similarity 100.0%; Pred No. 1.2e-136;  
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	G G G R R D G G P A C G G F D L Y F I L D K S G V L H W N E I Y Y F V E Q L A H K F I S P Q L R M S F I V F S	60
Db	27	G G G R R D G G P A C G G F D L Y F I L D K S G V L H W N E I Y Y F V E Q L A H K F I S P Q L R M S F I V F S	86
Qy	61	T R G T T L M K L T E D R E Q T R G L E E L Q K V L P G G D T Y M H G F E R A S E Q I Y Y E N R Q G Y R T A S V I I	120
Db	87	T R G T T L M K L T E D R E Q T R G L E E L Q K V L P G G D T Y M H G F E R A S E Q I Y Y E N R Q G Y R T A S V I I	146
Qy	121	A L T D G E L H E D L F P Y S R E A N R S R D L G A I V C V G K V D F N E T Q L A R I A D S K D H V F P V N D G F Q	180
Db	147	A L T D G E L H E D L F P Y S R E A N R S R D L G A I V C V G K V D F N E T Q L A R I A D S K D H V F P V N D G F Q	206
Qy	181	A L Q G I I H S I L K K S C I E I L A A E P S T I C A G E S F Q V V V R G N G F R H A R N V D R V L C S F K I N D S V T	240
Db	207	A L Q G I I H S I L K K S C I E I L A A E P S T I C A G E S F Q V V V R G N G F R H A R N V D R V L C S F K I N D S V T	266
Qy	241	L N E K P F S V E D T Y L L C C P A P I L K E V G M K A A L Q V S M N D G L S F I S S S V I T T H C S D G S	295
Db	267	L N E K P F S V E D T Y L L C C P A P I L K E V G M K A A L Q V S M N D G L S F I S S S V I T T H C S D G S	321

US-60-389-987-1823

Query Match 100.0%; Score 1536; DB 7; Length 564;  
Best Local Similarity 100.0%; Pred. No. 1.2e-136; Indels 0; Gaps 0;  
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGGRRDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 60  
DB 27 GGGRRDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 86  
QY 61 TRGTTLMKLTEDREQIROGLELQKVLPGDYMHEGFERASEQIYYENROGYRTASVII 120  
DB 87 TRGTTLMKLTEDREQIROGLELQKVLPGDYMHEGFERASEQIYYENROGYRTASVII 146  
QY 121 ALTDGELHEDLFFYSERANRDLGAIYCVGVKDFNETQLARIADSKDHVPVNDGQF 180  
DB 147 ALTDGELHEDLFFYSERANRDLGAIYCVGVKDFNETQLARIADSKDHVPVNDGQF 206  
QY 181 ALOGIIHSILKKSCIEILAAEPSTICAGESFQVYVGRNGFRHARNVDRVLCFSKINDSVT 240  
DB 207 ALOGIIHSILKKSCIEILAAEPSTICAGESFQVYVGRNGFRHARNVDRVLCFSKINDSVT 266  
QY 241 LNEKPFVSDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSVIITTHCSDGS 295  
DB 267 LNEKPFVSDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSVIITTHCSDGS 321

RESULT 4  
US-10-038-307-18  
; Sequence 18, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18  
; LENGTH: 551  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-038-307-18

Query Match 99.7%; Score 1532; DB 6; Length 551;  
Best Local Similarity 100.0%; Pred. No. 2.7e-136; Indels 0; Gaps 0;  
Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGGRRDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 60  
DB 27 GGGRRDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 86  
QY 61 TRGTTLMKLTEDREQIROGLELQKVLPGDYMHEGFERASEQIYYENROGYRTASVII 120  
DB 87 TRGTTLMKLTEDREQIROGLELQKVLPGDYMHEGFERASEQIYYENROGYRTASVII 146  
QY 121 ALTDGELHEDLFFYSERANRDLGAIYCVGVKDFNETQLARIADSKDHVPVNDGQF 180  
DB 147 ALTDGELHEDLFFYSERANRDLGAIYCVGVKDFNETQLARIADSKDHVPVNDGQF 206  
QY 181 ALOGIIHSILKKSCIEILAAEPSTICAGESFQVYVGRNGFRHARNVDRVLCFSKINDSVT 240  
DB 207 ALOGIIHSILKKSCIEILAAEPSTICAGESFQVYVGRNGFRHARNVDRVLCFSKINDSVT 266  
QY 241 LNEKPFVSDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSVIITTHCSDG 294  
DB 267 LNEKPFVSDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSVIITTHCSDG 320

US-60-389-987-1823

Query Match 100.0%; Score 1536; DB 1; Length 564;  
Best Local Similarity 100.0%; Pred. No. 1.2e-136; Indels 0; Gaps 0;  
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGGRRDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 60  
DB 27 GGGRRDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 86  
QY 61 TRGTTLMKLTEDREQIROGLELQKVLPGDYMHEGFERASEQIYYENROGYRTASVII 120  
DB 87 TRGTTLMKLTEDREQIROGLELQKVLPGDYMHEGFERASEQIYYENROGYRTASVII 146  
QY 121 ALTDGELHEDLFFYSERANRDLGAIYCVGVKDFNETQLARIADSKDHVPVNDGQF 180  
DB 147 ALTDGELHEDLFFYSERANRDLGAIYCVGVKDFNETQLARIADSKDHVPVNDGQF 206  
QY 181 ALOGIIHSILKKSCIEILAAEPSTICAGESFQVYVGRNGFRHARNVDRVLCFSKINDSVT 240  
DB 207 ALOGIIHSILKKSCIEILAAEPSTICAGESFQVYVGRNGFRHARNVDRVLCFSKINDSVT 266  
QY 241 LNEKPFVSDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSVIITTHCSDGS 295  
DB 267 LNEKPFVSDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSVIITTHCSDGS 321

RESULT 3  
US-60-389-987-1823  
; Sequence 1823, Application US/60389987  
; GENERAL INFORMATION:  
; APPLICANT: Ghosh, Soumitra S.  
; APPLICANT: Fahy, Eoin D.  
; APPLICANT: Zhang, Bing  
; APPLICANT: Gibson, Bradford W.  
; APPLICANT: Taylor, Steven W.  
; APPLICANT: Glenn, Gary M.  
; APPLICANT: Warnock, Dale E.  
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
; FILE REFERENCE: 660088.465P2  
; CURRENT APPLICATION NUMBER: US/60/389,987  
; CURRENT FILING DATE: 2002-06-17  
; NUMBER OF SEQ ID NOS: 3025  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1823  
; LENGTH: 564  
; TYPE: PRT  
; ORGANISM: Homo sapiens

Query Match 100.0%; Score 1536; DB 1; Length 564;  
Best Local Similarity 100.0%; Pred. No. 1.2e-136; Indels 0; Gaps 0;  
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGGRRDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 60  
DB 27 GGGRRDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 86  
QY 61 TRGTTLMKLTEDREQIROGLELQKVLPGDYMHEGFERASEQIYYENROGYRTASVII 120  
DB 87 TRGTTLMKLTEDREQIROGLELQKVLPGDYMHEGFERASEQIYYENROGYRTASVII 146  
QY 121 ALTDGELHEDLFFYSERANRDLGAIYCVGVKDFNETQLARIADSKDHVPVNDGQF 180  
DB 147 ALTDGELHEDLFFYSERANRDLGAIYCVGVKDFNETQLARIADSKDHVPVNDGQF 206  
QY 181 ALOGIIHSILKKSCIEILAAEPSTICAGESFQVYVGRNGFRHARNVDRVLCFSKINDSVT 240  
DB 207 ALOGIIHSILKKSCIEILAAEPSTICAGESFQVYVGRNGFRHARNVDRVLCFSKINDSVT 266  
QY 241 LNEKPFVSDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSVIITTHCSDGS 295  
DB 267 LNEKPFVSDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSVIITTHCSDGS 321

US-60-389-987-1823

```
RESULT 5
US-10-038-307-2
; Sequence 2, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-2

Query Match 99.0%; Score 1520; DB 6; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.8e-135;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGRRDGGPACYGCGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 60
Db 27 GGGRRDGGPACYGCGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 86
Qy 61 TRGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGGERASEQIYYENRQGYRTASVII 120
Db 27 GGGRRDGGPACYGCGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 86
Qy 87 TRGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGGERASEQIYYENRQGYRTASVII 146
Db 87 TRGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGGERASEQIYYENRQGYRTASVII 146
Qy 121 ALTDGELHEDLFFYSERANRSDLGAIYCVGVKDFNETQLARIADSKDHVFPVNDGFG 180
Db 147 ALTDGELHEDLFFYSERANRSDLGAIYCVGVKDFNETQLARIADSKDHVFPVNDGFG 206
Qy 181 ALQGIHSLKKSCIEILAAEPSTICAGESFOVVYRGNGFRHARNVDRVLCSEFKINDSV 240
Db 207 ALQGIHSLKKSCIEILAAEPSTICAGESFOVVYRGNGFRHARNVDRVLCSEFKINDSV 266
Qy 241 LNEKPFSEVEDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSSVIITTHCS 292
Db 267 LNEKPFSEVEDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSSVIITTHCS 318

RESULT 6
US-10-038-307-24
; Sequence 24, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-24

Query Match 99.0%; Score 1520; DB 6; Length 345;
Best Local Similarity 100.0%; Pred. No. 1.9e-135;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGRRDGGPACYGCGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 60
```

```
Db 27 GGGRRDGGPACYGCGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 86
Qy 61 TRGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGGERASEQIYYENRQGYRTASVII 120
Db 87 TRGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGGERASEQIYYENRQGYRTASVII 146
Qy 121 ALTDGELHEDLFFYSERANRSDLGAIYCVGVKDFNETQLARIADSKDHVFPVNDGFG 180
Db 147 ALTDGELHEDLFFYSERANRSDLGAIYCVGVKDFNETQLARIADSKDHVFPVNDGFG 206
Qy 181 ALQGIHSLKKSCIEILAAEPSTICAGESFOVVYRGNGFRHARNVDRVLCSEFKINDSV 240
Db 207 ALQGIHSLKKSCIEILAAEPSTICAGESFOVVYRGNGFRHARNVDRVLCSEFKINDSV 266
Qy 241 LNEKPFSEVEDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSSVIITTHCS 292
Db 267 LNEKPFSEVEDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSSVIITTHCS 318

RESULT 7
US-10-038-307-20
; Sequence 20, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-20

Query Match 99.0%; Score 1520; DB 6; Length 564;
Best Local Similarity 100.0%; Pred. No. 3.8e-135;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGRRDGGPACYGCGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 60
Db 27 GGGRRDGGPACYGCGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 86
Qy 61 TRGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGGERASEQIYYENRQGYRTASVII 120
Db 87 TRGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGGERASEQIYYENRQGYRTASVII 146
Qy 121 ALTDGELHEDLFFYSERANRSDLGAIYCVGVKDFNETQLARIADSKDHVFPVNDGFG 180
Db 147 ALTDGELHEDLFFYSERANRSDLGAIYCVGVKDFNETQLARIADSKDHVFPVNDGFG 206
Qy 181 ALQGIHSLKKSCIEILAAEPSTICAGESFOVVYRGNGFRHARNVDRVLCSEFKINDSV 240
Db 207 ALQGIHSLKKSCIEILAAEPSTICAGESFOVVYRGNGFRHARNVDRVLCSEFKINDSV 266
Qy 241 LNEKPFSEVEDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSSVIITTHCS 292
Db 267 LNEKPFSEVEDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSSVIITTHCS 318
```

```
RESULT 8
PCT-US02-08253-194
; Sequence 194, Application PC/TUS0208253
; GENERAL INFORMATION:
; APPLICANT: Carson-Walter, Eleanor
; APPLICANT: St. Croix, Brad
; APPLICANT: Vogelstein, Bert

Query Match 99.0%; Score 1520; DB 6; Length 345;
Best Local Similarity 100.0%; Pred. No. 1.9e-135;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGRRDGGPACYGCGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 60
Db 27 GGGRRDGGPACYGCGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 86
Qy 61 TRGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGGERASEQIYYENRQGYRTASVII 120
Db 87 TRGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGGERASEQIYYENRQGYRTASVII 146
Qy 121 ALTDGELHEDLFFYSERANRSDLGAIYCVGVKDFNETQLARIADSKDHVFPVNDGFG 180
Db 147 ALTDGELHEDLFFYSERANRSDLGAIYCVGVKDFNETQLARIADSKDHVFPVNDGFG 206
Qy 181 ALQGIHSLKKSCIEILAAEPSTICAGESFOVVYRGNGFRHARNVDRVLCSEFKINDSV 240
Db 207 ALQGIHSLKKSCIEILAAEPSTICAGESFOVVYRGNGFRHARNVDRVLCSEFKINDSV 266
Qy 241 LNEKPFSEVEDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSSVIITTHCS 292
Db 267 LNEKPFSEVEDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSSVIITTHCS 318
```

APPLICANT: Kinzler, Kenneth  
TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
FILE REFERENCE: 1107.00179  
CURRENT APPLICATION NUMBER: PCT/US02/08253  
CURRENT FILING DATE: 2002-04-10  
PRIOR APPLICATION NUMBER: 60/282,850  
PRIOR FILING DATE: 2001-04-11  
PRIOR APPLICATION NUMBER: 60/308,829  
PRIOR FILING DATE: 2001-08-01  
NUMBER OF SEQ ID NOS: 359  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 194  
LENGTH: 562  
TYPE: PRT  
ORGANISM: Mus musculus  
PCT-US02-08253-194

Query Match 98.9%; Score 1519; DB 1; Length 562;  
Best Local Similarity 98.6%; Pred. No. 4.7e-135;  
Matches 291; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGRRDGGPACYGCGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 60  
Db 25 GHGRRDGGPACYGCGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 84  
QY 61 TRGTTLMKLTEDREQIROGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 120  
Db 85 TRGTTLMKLTEDREQIROGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 144  
QY 121 ALTDGELHEDLFFYSREANRSDLGAIYVCVGVKDFNETQLARIADSKDHVFPVNDGFQ 180  
Db 145 ALTDGELHEDLFFYSREANRSDLGAIYVCVGVKDFNETQLARIADSKDHVFPVNDGFQ 204  
QY 181 ALQGIHHSILKKSCIEILAAEPSTICAGESFQVVVGRNGFRHARNVDRVLCSEKINDSVT 240  
Db 205 ALQGIHHSILKKSCIEILAAEPSTICAGESFQVVVGRNGFRHARNVDRVLCSEKINDSVT 264  
QY 241 LNEKPFSEVEDTYLLCPAPILKEVGKMAALQVSMNDGLSFSSSVITTTTHCSGDS 295  
Db 265 LNEKPFSEVEDTYLLCPAPILKEVGKMAALQVSMNDGLSFSSSVITTTTHCSGDS 319

RESULT 10  
US-10-038-307-26  
; Sequence 26, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 26  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-038-307-26

Query Match 98.1%; Score 1507; DB 6; Length 328;  
Best Local Similarity 100.0%; Pred. No. 3e-134;  
Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRDGGPACYGCGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 60  
Db 27 GGGRRDGGPACYGCGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 86  
QY 61 TRGTTLMKLTEDREQIROGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 120  
Db 87 TRGTTLMKLTEDREQIROGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 146  
QY 121 ALTDGELHEDLFFYSREANRSDLGAIYVCVGVKDFNETQLARIADSKDHVFPVNDGFQ 180  
Db 147 ALTDGELHEDLFFYSREANRSDLGAIYVCVGVKDFNETQLARIADSKDHVFPVNDGFQ 206  
QY 181 ALQGIHHSILKKSCIEILAAEPSTICAGESFQVVVGRNGFRHARNVDRVLCSEKINDSVT 240  
Db 207 ALQGIHHSILKKSCIEILAAEPSTICAGESFQVVVGRNGFRHARNVDRVLCSEKINDSVT 266  
QY 241 LNEKPFSEVEDTYLLCPAPILKEVGKMAALQVSMNDGLSFSSSVITTTTH 290  
Db 267 LNEKPFSEVEDTYLLCPAPILKEVGKMAALQVSMNDGLSFSSSVITTTTH 316

RESULT 11  
US-10-038-307-22  
; Sequence 22, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK

APPLICANT: Kinzler, Kenneth  
TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
FILE REFERENCE: 1107.00179  
CURRENT APPLICATION NUMBER: PCT/US02/08253  
CURRENT FILING DATE: 2002-04-10  
PRIOR APPLICATION NUMBER: 60/282,850  
PRIOR FILING DATE: 2001-04-11  
PRIOR APPLICATION NUMBER: 60/308,829  
PRIOR FILING DATE: 2001-08-01  
NUMBER OF SEQ ID NOS: 359  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 194  
LENGTH: 562  
TYPE: PRT  
ORGANISM: Mus musculus  
PCT-US02-08253-194

Query Match 98.9%; Score 1519; DB 1; Length 562;  
Best Local Similarity 98.6%; Pred. No. 4.7e-135;  
Matches 291; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGRRDGGPACYGCGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 60  
Db 25 GHGRRDGGPACYGCGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 84  
QY 61 TRGTTLMKLTEDREQIROGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 120  
Db 85 TRGTTLMKLTEDREQIROGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 144  
QY 121 ALTDGELHEDLFFYSREANRSDLGAIYVCVGVKDFNETQLARIADSKDHVFPVNDGFQ 180  
Db 145 ALTDGELHEDLFFYSREANRSDLGAIYVCVGVKDFNETQLARIADSKDHVFPVNDGFQ 204  
QY 181 ALQGIHHSILKKSCIEILAAEPSTICAGESFQVVVGRNGFRHARNVDRVLCSEKINDSVT 240  
Db 205 ALQGIHHSILKKSCIEILAAEPSTICAGESFQVVVGRNGFRHARNVDRVLCSEKINDSVT 264  
QY 241 LNEKPFSEVEDTYLLCPAPILKEVGKMAALQVSMNDGLSFSSSVITTTTHCSGDS 295  
Db 265 LNEKPFSEVEDTYLLCPAPILKEVGKMAALQVSMNDGLSFSSSVITTTTHCSGDS 319

RESULT 9  
PCT-US02-08253-301  
; Sequence 301, Application PC/TUS0208253  
; GENERAL INFORMATION:  
; APPLICANT: Carson-Walter, Eleanor  
; APPLICANT: St. Croix, Brad  
; APPLICANT: Vogelstein, Bert  
; APPLICANT: Kinzler, Kenneth  
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
; FILE REFERENCE: 1107.00179  
; CURRENT APPLICATION NUMBER: PCT/US02/08253  
; CURRENT FILING DATE: 2002-04-10  
; PRIOR APPLICATION NUMBER: 60/282,850  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/308,829  
; PRIOR FILING DATE: 2001-08-01  
; NUMBER OF SEQ ID NOS: 359  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 301  
; LENGTH: 562  
; TYPE: PRT  
; ORGANISM: Mouse  
PCT-US02-08253-301

Query Match 98.9%; Score 1519; DB 1; Length 562;  
Best Local Similarity 98.6%; Pred. No. 4.7e-135;  
Matches 291; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGRRDGGPACYGCGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 60

```

Db      87  TRGTTLMKLTREDRIQIGLEEIQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 146
QY      121  ALTDGELHEDLFFYSEREANRSDLGAIYVCVGKDFNETQLARIADSKDHVPVNDGFG 180
Db      147  ALTDGELHEDLFFYS--EANRSDLGAIYVCVGKDFNETQLARIADSKDHVPVNDGFG 204
QY      181  ALQGIHSILKKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSPKINDSVT 240
Db      205  ALQGIHSILKKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSPKINDSVT 264
QY      241  LNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVITTHCSDGS 295
Db      265  LNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVITTHCSDGS 319

RESULT 13
US-10-038-307-16
; Sequence 16, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-16

Query Match 97.7%; Score 1500; DB 6; Length 543;
Best Local Similarity 98.6%; Pred. No. 2.8e-133;
Matches 291; Conservative 0; Mismatches 2; Indels 2; Gaps

QY      1  GGGRRDDGGPACVGGPDLYFLDKSGSVLHHWNEIYFVEQLAHKFI SPOLRMSFIVFS 60
Db      27  GGGRRDDGGPACVGGPDLYFLDKSGSVLHHWNEIYFVEQLAHKFI SPOLRMSFIVFS 86
QY      61  TRGTTLMKLTREDRIQIGLEEIQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 120
Db      87  TRGTTLMKLTREDRIQIGLEEIQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 146
QY      121  ALTDGELHEDLFFYSEREANRSDLGAIYVCVGKDFNETQLARIADSKDHVPVNDGFG 180
Db      147  ALTDGELHEDLFFYS--EANRSDLGAIYVCVGKDFNETQLARIADSKDHVPVNDGFG 204
QY      181  ALQGIHSILKKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSPKINDSVT 240
Db      205  ALQGIHSILKKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSPKINDSVT 264
QY      241  LNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVITTHCSDGS 295
Db      265  LNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVITTHCSDGS 319

RESULT 14
US-10-038-307-10
; Sequence 10, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307

```

Db 198 QGIIHSILKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSEKINDSVTLN 257  
QY 243 EKPFSEVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVITTTTHGSDGS 295  
Db 258 EKPFSEVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVITTTTHSSPKS 310

Search completed: August 9, 2002, 10:35:55  
Job time: 353 sec

Query Match 97.0%; Score 1490; DB 6; Length 543;  
Best Local Similarity 98.3%; Pred. No. 2.5e-132; Indels 2; Gaps 1;  
Matches 290; Conservative 0; Mismatches 3;

QY 1 GGGREDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 60  
Db 27 GGGREDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 86  
QY 61 TRGTTLMKLTEDREQIRGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 120  
Db 87 TRGTTLMKLTEDREQIRGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 146  
QY 121 ALTDGELHEDLFFYSERANRDLGAIYVCVKDFNETQLARIADSKDHVPVNDGFQ 180  
Db 147 ALTDGELHEDLFFYS--EANRSDLGAIYVCVKDFNETQLARIADSKDHVPVNDGFQ 204  
QY 181 ALQGIHSILKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSEKINDSVT 240  
Db 205 ALQGIHSILKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSEKINDSVT 264  
QY 241 LNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVITTTTHGSDGS 295  
Db 265 LNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVITTTTHSSPKS 319

RESULT 15  
US-10-038-307-12  
; Sequence 12, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 534  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-038-307-12

Query Match 96.3%; Score 1479; DB 6; Length 534;  
Best Local Similarity 98.3%; Pred. No. 2.6e-131;  
Matches 288; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

QY 3 GGGREDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFSR 62  
Db 20 GGGREDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFSR 79  
QY 63 GTTLMKLTEDREQIRGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVIIAL 122  
Db 80 GTTLMKLTEDREQIRGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVIIAL 139  
QY 123 TDGELHEDLFFYSERANRDLGAIYVCVKDFNETQLARIADSKDHVPVNDGFQAL 182  
Db 140 TDGELHEDLFFYS--EANRSDLGAIYVCVKDFNETQLARIADSKDHVPVNDGFQAL 197  
QY 183 QGIIHSILKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSEKINDSVTLN 242

---

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.  
OM protein - protein search, using sw model  
Run on: August 9, 2002, 10:44:29 ; Search time 507.33 seconds  
(without alignments)  
204.668 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_27\_321  
Perfect score: 1536  
Sequence: 1 GGGRRDGGPACYGFDLY.....GLSFSSSVIITTHCSDGS 295

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues  
Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/paa/US080\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep.\*  
7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep.\*  
8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep.\*  
9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep.\*  
10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep.\*  
11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep.\*  
12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep.\*  
13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep.\*  
14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep.\*  
15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep.\*  
16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep.\*  
17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep.\*  
18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep.\*  
19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep.\*  
20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep.\*  
21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep.\*  
22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep.\*  
23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep.\*  
24: /cgn2\_6/ptodata/2/paa/US100\_COMB.pep.\*  
25: /cgn2\_6/ptodata/2/paa/US101\_COMB.pep.\*  
26: /cgn2\_6/ptodata/2/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1536	100.0	368	23	US-09-970-076-2
2	1536	100.0	403	1	PCT-US00-30045-94
3	1536	100.0	403	1	PCT-US01-11988-621
4	1536	100.0	403	22	US-09-833-245-621
5	1536	100.0	564	23	US-09-918-715-187
6	1536	100.0	564	23	US-09-918-715-232
7	1536	100.0	564	23	US-09-970-076-6

ALIGNMENTS

RESULT 1

US-09-970-076-2

; Sequence 2, Application US/09970076

; GENERAL INFORMATION:

; APPLICANT: Young, John A.T.

; APPLICANT: Bradley, Kenneth A.

; APPLICANT: Collier, Robert J.

; APPLICANT: Mogridge, Jeremy S.

; TITLE OF INVENTION: Anthrax Toxin Receptor

; FILE REFERENCE: 960296.97745

; CURRENT APPLICATION NUMBER: US/09/970,076

; PRIOR FILING DATE: 2001-10-03

; PRIOR APPLICATION NUMBER: 60/251,481

; PRIOR FILING DATE: 2000-12-05

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 368

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-970-076-2

Query Match 100.0%; Score 1536; DB 23; Length 368;  
Best Local Similarity 100.0%; Pred. No. 6.5e-151;  
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGRRDGGPACYGFDLYFLDKSGVSLHHWNHNYFVEQLAHKFISPOLRMSFVFS 60  
|||||  
Db 27 GGGRRDGGPACYGFDLYFLDKSGVSLHHWNHNYFVEQLAHKFISPOLRMSFVFS 86

Sequence 125, App  
Sequence 620, App  
Sequence 620, App  
Sequence 9, Appli  
Sequence 9, Appli  
Sequence 9, Appli  
Sequence 12, Appli  
Sequence 8, Appli  
Sequence 194, App  
Sequence 301, App  
Sequence 51, Appl  
Sequence 30, Appl  
Sequence 51, Appl  
Sequence 51, Appl  
Sequence 132, App  
Sequence 132, App  
Sequence 50, Appl  
Sequence 2121, Ap  
Sequence 10, Appl  
Sequence 105, App  
Sequence 105, App  
Sequence 59, Appl  
Sequence 59, Appl  
Sequence 107, App  
Sequence 107, App  
Sequence 2, Appli  
Sequence 57, Appl  
Sequence 57, Appl  
Sequence 52, Appl  
Sequence 10, Appl  
Sequence 20, Appl  
Sequence 6, Appli  
Sequence 10, Appl  
Sequence 6, Appli  
Sequence 16, Appl

Fri Aug 9 10:57:02 2002

```

; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: 60/229, 358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256, 931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199, 384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 621
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-11988-621

Query Match 100.0%; Score 1536; DB 1; Length 403;
Best Local Similarity 100.0%; Pred. No. 7.4e-151;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRDDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 60
Db 27 GGGRRDDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 86
QY 61 TRGTTLMKLTEDREQIROGLELQKVLPGDYMHEGFERASEQIYYENRQGYRTASVII 120
Db 87 TRGTTLMKLTEDREQIROGLELQKVLPGDYMHEGFERASEQIYYENRQGYRTASVII 146
QY 121 ALTDGELHEDLFFYSREANRSDLGAIYVCVKDFNETQIARIADSKDHVFPVNDGFG 180
Db 147 ALTDGELHEDLFFYSREANRSDLGAIYVCVKDFNETQIARIADSKDHVFPVNDGFG 206
QY 181 ALQGIHSLKSCIEILAAEPSTICAGESFOVVYRNGNGFRHARNVDRVLCFSKINDSVT 240
Db 207 ALQGIHSLKSCIEILAAEPSTICAGESFOVVYRNGNGFRHARNVDRVLCFSKINDSVT 266
QY 241 LNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVIITTHCSDGS 295
Db 267 LNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVIITTHCSDGS 321

RESULT 4
US-09-833-245-621
; Sequence 621, Application US/09833245
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229, 358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256, 931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199, 384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 621
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-245-621

Query Match 100.0%; Score 1536; DB 22; Length 403;
Best Local Similarity 100.0%; Pred. No. 7.4e-151;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRDDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 60
Db 27 GGGRRDDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 86

PCT-US00-30045-94
; Sequence 94, Application PC/TUS0030045
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: PS708PCT
; CURRENT APPLICATION NUMBER: PCT/US00/30045
; CURRENT FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/163,581
; PRIOR FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: 60/215,133
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 201
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 94
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US00-30045-94

Query Match 100.0%; Score 1536; DB 1; Length 403;
Best Local Similarity 100.0%; Pred. No. 7.4e-151;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRDDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 60
Db 27 GGGRRDDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 86
QY 61 TRGTTLMKLTEDREQIROGLELQKVLPGDYMHEGFERASEQIYYENRQGYRTASVII 120
Db 87 TRGTTLMKLTEDREQIROGLELQKVLPGDYMHEGFERASEQIYYENRQGYRTASVII 146
QY 121 ALTDGELHEDLFFYSREANRSDLGAIYVCVKDFNETQIARIADSKDHVFPVNDGFG 180
Db 147 ALTDGELHEDLFFYSREANRSDLGAIYVCVKDFNETQIARIADSKDHVFPVNDGFG 206
QY 181 ALQGIHSLKSCIEILAAEPSTICAGESFOVVYRNGNGFRHARNVDRVLCFSKINDSVT 240
Db 207 ALQGIHSLKSCIEILAAEPSTICAGESFOVVYRNGNGFRHARNVDRVLCFSKINDSVT 266
QY 241 LNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVIITTHCSDGS 295
Db 267 LNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVIITTHCSDGS 321

RESULT 3
PCT-US01-11988-621
; Sequence 621, Application PC/TUS0111988
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: PCT/US01/11988
```

QY 61 TRGTTLMKLTEDREQIRGLEELQKLVPGDGYMHGFERASEQIYYENRQGYRTASVII 120  
Db 87 TRGTTLMKLTEDREQIRGLEELQKLVPGDGYMHGFERASEQIYYENRQGYRTASVII 146  
QY 121 ALTDGELHEDLFFYSERANRSDLGAIYVCVGVKDFNETQLARIADSKDHVFPVNDGQ 180  
Db 147 ALTDGELHEDLFFYSERANRSDLGAIYVCVGVKDFNETQLARIADSKDHVFPVNDGQ 206  
QY 181 ALQGIHSLKKSCIEILAAEPSTICAGESFOVVRGNGFRHARNVDRVLCFKINDSVT 240  
Db 207 ALQGIHSLKKSCIEILAAEPSTICAGESFOVVRGNGFRHARNVDRVLCFKINDSVT 266  
QY 241 LNEKPFVSDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSSVIITTHCSDGS 295  
Db 267 LNEKPFVSDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSSVIITTHCSDGS 321

## RESULT 5

US-09-918-715-187  
; Sequence 187, Application US/09918715

; GENERAL INFORMATION:

; APPLICANT: Brad St. Croix

; APPLICANT: Bert Vogelstein

; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS

; FILE REFERENCE: 1107.00134

; CURRENT APPLICATION NUMBER: US/09/918,715

; CURRENT FILING DATE: 2001-08-01

; PRIOR APPLICATION NUMBER: 60/222,599

; PRIOR FILING DATE: 2000-08-02

; PRIOR APPLICATION NUMBER: 60/224,360

; PRIOR FILING DATE: 2000-08-11

; PRIOR APPLICATION NUMBER: 60/282,850

; PRIOR FILING DATE: 2000-04-11

; NUMBER OF SEQ ID NOS: 358

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 187

; LENGTH: 564

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-918-715-187

Query Match 100.0%; Score 1536; DB 23; Length 564;  
Best Local Similarity 100.0%; Pred. No. 1.2e-150;  
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRDGGPACYGFDLYFLDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVES 60  
Db 27 GGGRRDGGPACYGFDLYFLDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVES 86  
QY 61 TRGTTLMKLTEDREQIRGLEELQKLVPGDGYMHGFERASEQIYYENRQGYRTASVII 120  
Db 87 TRGTTLMKLTEDREQIRGLEELQKLVPGDGYMHGFERASEQIYYENRQGYRTASVII 146  
QY 121 ALTDGELHEDLFFYSERANRSDLGAIYVCVGVKDFNETQLARIADSKDHVFPVNDGQ 180  
Db 147 ALTDGELHEDLFFYSERANRSDLGAIYVCVGVKDFNETQLARIADSKDHVFPVNDGQ 206  
QY 181 ALQGIHSLKKSCIEILAAEPSTICAGESFOVVRGNGFRHARNVDRVLCFKINDSVT 240  
Db 207 ALQGIHSLKKSCIEILAAEPSTICAGESFOVVRGNGFRHARNVDRVLCFKINDSVT 266  
QY 241 LNEKPFVSDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSSVIITTHCSDGS 295  
Db 267 LNEKPFVSDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSSVIITTHCSDGS 321

## RESULT 6

US-09-918-715-232

; Sequence 232, Application US/09918715

; GENERAL INFORMATION:

; APPLICANT: Brad St. Croix

Query Match 100.0%; Score 1536; DB 23; Length 564;  
Best Local Similarity 100.0%; Pred. No. 1.2e-150;  
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; APPLICANT: Bert Vogelstein  
; APPLICANT: Kenneth Kinzler  
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
; FILE REFERENCE: 1107.00134  
; CURRENT APPLICATION NUMBER: US/09/918,715  
; CURRENT FILING DATE: 2001-08-01  
; PRIOR APPLICATION NUMBER: 60/222,599  
; PRIOR FILING DATE: 2000-08-02  
; PRIOR APPLICATION NUMBER: 60/224,360  
; PRIOR FILING DATE: 2000-08-11  
; PRIOR APPLICATION NUMBER: 60/282,850  
; PRIOR FILING DATE: 2000-04-11  
; NUMBER OF SEQ ID NOS: 358  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 232  
; LENGTH: 564  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-918-715-232

Query Match 100.0%; Score 1536; DB 23; Length 564;  
Best Local Similarity 100.0%; Pred. No. 1.2e-150;  
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRDGGPACYGFDLYFLDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVES 60  
Db 27 GGGRRDGGPACYGFDLYFLDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVES 86  
QY 61 TRGTTLMKLTEDREQIRGLEELQKLVPGDGYMHGFERASEQIYYENRQGYRTASVII 120  
Db 87 TRGTTLMKLTEDREQIRGLEELQKLVPGDGYMHGFERASEQIYYENRQGYRTASVII 146  
QY 121 ALTDGELHEDLFFYSERANRSDLGAIYVCVGVKDFNETQLARIADSKDHVFPVNDGQ 180  
Db 147 ALTDGELHEDLFFYSERANRSDLGAIYVCVGVKDFNETQLARIADSKDHVFPVNDGQ 206  
QY 181 ALQGIHSLKKSCIEILAAEPSTICAGESFOVVRGNGFRHARNVDRVLCFKINDSVT 240  
Db 207 ALQGIHSLKKSCIEILAAEPSTICAGESFOVVRGNGFRHARNVDRVLCFKINDSVT 266  
QY 241 LNEKPFVSDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSSVIITTHCSDGS 295  
Db 267 LNEKPFVSDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSSVIITTHCSDGS 321

## RESULT 7

US-09-970-076-6  
; Sequence 6, Application US/09970076

; GENERAL INFORMATION:

; APPLICANT: Young, John A.T.

; APPLICANT: Bradley, Kenneth A.

; APPLICANT: Collier, Robert J.

; APPLICANT: Mogridge, Jeremy S.

; TITLE OF INVENTION: Anthrax Toxin Receptor

; FILE REFERENCE: 960296.97745

; CURRENT APPLICATION NUMBER: US/09/970,076

; CURRENT FILING DATE: 2001-10-03

; PRIOR APPLICATION NUMBER: 60/251,481

; PRIOR FILING DATE: 2000-12-05

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 6

; LENGTH: 564

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-970-076-6

Fri Aug 9 10:57:02 2002

```

QY 1 GGGRRDGGPACGGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVS 60
|
|
|
Db 27 GGGRRDGGPACGGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVS 86
|
|
|
QY 61 TRGTTMLKLTEDREQIROGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 120
|
|
|
Db 87 TRGTTMLKLTEDREQIROGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 146
|
|
|
QY 121 ALTDGELHEDLFFYSEREANRDLGAIVYCVGVKDFNETQLARIADSKDHVPVNDGFG 180
|
|
|
Db 147 ALTDGELHEDLFFYSEREANRDLGAIVYCVGVKDFNETQLARIADSKDHVPVNDGFG 206
|
|
|
QY 181 ALQGIHSILKSKCIEILAAEPSTICAGESFQVYVGRNGFRHARNVDRVLCSEKINDSVT 240
|
|
|
Db 207 ALQGIHSILKSKCIEILAAEPSTICAGESFQVYVGRNGFRHARNVDRVLCSEKINDSVT 266
|
|
|
QY 241 LNEKPFVSDTYLLCPAPILKEVGKMAALQVSMNDGLSFISSSVITTTTHCSDGS 295
|
|
|
Db 267 LNEKPFVSDTYLLCPAPILKEVGKMAALQVSMNDGLSFISSSVITTTTHCSDGS 321
|
|
|
RESULT 8
PCT-US00-30045-125
; Sequence 125, Application PC/TUS0030045
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: PS708PCT
; CURRENT APPLICATION NUMBER: PCT/US00/30045
; CURRENT FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/163,581
; PRIOR FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: 60/215,133
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 201
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 125
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (175)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (320)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (331)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (368)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
PCT-US00-30045-125

Query Match 99.2%; Score 1524; DB 1; Length 403;
Best Local Similarity 99.3%; Pred. No. 1.3e-149;
Matches 293; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGRRDGGPACGGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVS 60
|
|
|
Db 27 GGGRRDGGPACGGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVS 86
|
|
|
QY 61 TRGTTMLKLTEDREQIROGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 120
|
|
|
Db 87 TRGTTMLKLTEDREQIROGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 146
|
|
|
QY 121 ALTDGELHEDLFFYSEREANRDLGAIVYCVGVKDFNETQLARIADSKDHVPVNDGFG 180
|
|
|
Db 147 ALTDGELHEDLFFYSEREANRDLGAIVYCVGVKDFNETQLARIADSKDHVPVNDGFG 206
|
|
|
QY 181 ALQGIHSILKSKCIEILAAEPSTICAGESFQVYVGRNGFRHARNVDRVLCSEKINDSVT 240
|
|
|
Db 207 ALQGIHSILKSKCIEILAAEPSTICAGESFQVYVGRNGFRHARNVDRVLCSEKINDSVT 266
|
|
|
QY 241 LNEKPFVSDTYLLCPAPILKEVGKMAALQVSMNDGLSFISSSVITTTTHCSDGS 295
|
|
|
Db 267 LNEKPFVSDTYLLCPAPILKEVGKMAALQVSMNDGLSFISSSVITTTTHCSDGS 321
|
|
|
RESULT 10
US-09-833-245-620
; Sequence 620, Application US/09833245

```

```

|
|
|
Db 207 ALQGIHSILKSKCIEILAAEPSTICAGESFQVYVGRNGFRHARNVDRVLCSEKINDSVT 266
|
|
|
QY 241 LNEKPFVSDTYLLCPAPILKEVGKMAALQVSMNDGLSFISSSVITTTTHCSDGS 295
|
|
|
Db 267 LNEKPFVSDTYLLCPAPILKEVGKMAALQVSMNDGLSFISSSVITTTTHCSDGS 321
|
|
|
RESULT 9
PCT-US01-11988-620
; Sequence 620, Application PC/TUS0111988
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: PCT/US01/11988
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: 60/229, 358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256, 931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199, 384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 620
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (175)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (320)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (331)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (368)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
PCT-US01-11988-620

Query Match 99.2%; Score 1524; DB 1; Length 403;
Best Local Similarity 99.3%; Pred. No. 1.3e-149;
Matches 293; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGRRDGGPACGGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVS 60
|
|
|
Db 27 GGGRRDGGPACGGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVS 86
|
|
|
QY 61 TRGTTMLKLTEDREQIROGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 120
|
|
|
Db 87 TRGTTMLKLTEDREQIROGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 146
|
|
|
QY 121 ALTDGELHEDLFFYSEREANRDLGAIVYCVGVKDFNETQLARIADSKDHVPVNDGFG 180
|
|
|
Db 147 ALTDGELHEDLFFYSEREANRDLGAIVYCVGVKDFNETQLARIADSKDHVPVNDGFG 206
|
|
|
QY 181 ALQGIHSILKSKCIEILAAEPSTICAGESFQVYVGRNGFRHARNVDRVLCSEKINDSVT 240
|
|
|
Db 207 ALQGIHSILKSKCIEILAAEPSTICAGESFQVYVGRNGFRHARNVDRVLCSEKINDSVT 266
|
|
|
QY 241 LNEKPFVSDTYLLCPAPILKEVGKMAALQVSMNDGLSFISSSVITTTTHCSDGS 295
|
|
|
Db 267 LNEKPFVSDTYLLCPAPILKEVGKMAALQVSMNDGLSFISSSVITTTTHCSDGS 321
|
|
|
RESULT 10
US-09-833-245-620
; Sequence 620, Application US/09833245

```



Fri Aug 9 10:57:02 2002

```

RESULT 13
US-09-471-179-9
; Sequence 9, Application US/09471179
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 7853-173
; CURRENT APPLICATION NUMBER: US/09/471,179
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-471-179-9

Query Match          99.0%; Score 1520; DB 18; Length 333;
Best Local Similarity 100.0%; Pred. No. 2.6e-149;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRDDGGPACGCGFDLYFLDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 60
DB 27 GGGRRDDGGPACGCGFDLYFLDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 86

QY 61 TRGTTLMKLTEDREQIROGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 120
DB 87 TRGTTLMKLTEDREQIROGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 146

QY 121 ALTDGELHEDLFFYSERANRSRDGLAIYVCVKDFNETQLARIADSKDHVPVNDGFG 180
DB 147 ALTDGELHEDLFFYSERANRSRDGLAIYVCVKDFNETQLARIADSKDHVPVNDGFG 206

QY 181 ALQGIHSILKKSCIEILAAEPSTICAGESFQVVRGNGFRHARNVDRVLCSEFKINDSVT 240
DB 207 ALQGIHSILKKSCIEILAAEPSTICAGESFQVVRGNGFRHARNVDRVLCSEFKINDSVT 266

QY 241 LNEKPFVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVITTTTHCS 292
DB 267 LNEKPFVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVITTTTHCS 318

RESULT 14
US-09-599-596-9
; Sequence 9, Application US/09599596
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 09404/066001
; CURRENT APPLICATION NUMBER: US/09/599,596
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-599-596-9

Query Match          99.0%; Score 1520; DB 19; Length 333;
Best Local Similarity 100.0%; Pred. No. 2.6e-149;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRDDGGPACGCGFDLYFLDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 60
DB 27 GGGRRDDGGPACGCGFDLYFLDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFS 86

QY 61 TRGTTLMKLTEDREQIROGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 120
DB 87 TRGTTLMKLTEDREQIROGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVII 146

QY 121 ALTDGELHEDLFFYSERANRSRDGLAIYVCVKDFNETQLARIADSKDHVPVNDGFG 180
DB 147 ALTDGELHEDLFFYSERANRSRDGLAIYVCVKDFNETQLARIADSKDHVPVNDGFG 206

QY 181 ALQGIHSILKKSCIEILAAEPSTICAGESFQVVRGNGFRHARNVDRVLCSEFKINDSVT 240
DB 207 ALQGIHSILKKSCIEILAAEPSTICAGESFQVVRGNGFRHARNVDRVLCSEFKINDSVT 266

QY 241 LNEKPFVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVITTTTHCS 292
DB 267 LNEKPFVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVITTTTHCS 318

RESULT 15
US-09-796-753-12
; Sequence 12, Application US/09796753
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
```

; PRIOR FILING DATE: 2000-06-29  
; PRIOR APPLICATION NUMBER: 09/606,317  
; PRIOR FILING DATE: 2000-06-29  
; PRIOR APPLICATION NUMBER: 09/665,666  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: 09/677,751  
; PRIOR FILING DATE: 2000-09-30  
; NUMBER OF SEQ ID NOS: 162  
; SEQ ID NO 12  
; LENGTH: 333  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
us-09-796-753-12

Query Match 99.0%; Score 1520; DB 21; Length 333;  
Best Local Similarity 100.0%; Pred. No. 2.6e-149; Mismatches 0; Gaps 0;  
Matches 292; Conservative 0; Indels 0; Gaps 0;  
Qy 1 GGGRRDGGPACYGDFLYFLDKSGSVLHHWNEIYYVEQLAHKFISPOLRMSFIVFS 60  
Db 27 GGGRRDGGPACYGDFLYFLDKSGSVLHHWNEIYYVEQLAHKFISPOLRMSFIVFS 86  
Qy 61 TRGTTLMKLTEDREQIROGLEELQKVLPGDITYMHEGPERASEQIYYENROGYRTASVII 120  
Db 87 TRGTTLMKLTEDREQIROGLEELQKVLPGDITYMHEGPERASEQIYYENROGYRTASVII 146  
Qy 121 ALTGELHEDLFFYSERANKSRDLGATVYCVGVKDFNETOLARTADSKDHVPVNDGFO 180  
Db 147 ALTGELHEDLFFYSERANKSRDLGATVYCVGVKDFNETOLARTADSKDHVPVNDGFO 206  
Qy 181 ALQGIHSILKKSCTEILAAEPSTICAGESFQVVVGRNGFRHARNVDRVLCFSKINDSVT 240  
Db 207 ALQGIHSILKKSCTEILAAEPSTICAGESFQVVVGRNGFRHARNVDRVLCFSKINDSVT 266  
Qy 241 LNEKPFVSEDYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVITTTTHCS 292  
Db 267 LNEKPFVSEDYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVITTTTHCS 318

Search completed: August 9, 2002, 10:44:30  
Job time: 868 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:33:03 ; Search time 45.48 Seconds  
(without alignments)  
158.433 Million cell updates/sec

Title: us-09-970-076-2\_COPY\_27\_321

Perfect score: 1536

Sequence: 1 GCGGRRDGGPACYGFDLY.....GLSFSSSVITTHCSDGS 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCFUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	166	10.8	1155	1	US-08-286-889-46
2	166	10.8	1155	1	US-08-485-618-46
3	166	10.8	1155	1	US-08-362-652-46
4	166	10.8	1155	2	US-08-605-672-46
5	166	10.8	1155	2	US-08-482-293A-46
6	166	10.8	1155	2	US-08-943-363-46
7	166	10.8	1155	4	US-09-193-043-46
8	166	10.8	1161	1	US-08-485-618-53
9	166	10.8	1161	1	US-08-362-652-53
10	166	10.8	1161	2	US-08-605-672-53
11	166	10.8	1161	2	US-08-482-293A-53
12	166	10.8	1161	2	US-08-943-363-53
13	166	10.8	1161	4	US-09-193-043-53
14	159.5	10.4	1151	1	US-08-286-889-37
15	159.5	10.4	1151	1	US-08-485-618-37
16	159.5	10.4	1151	1	US-08-362-652-37
17	159.5	10.4	1151	2	US-08-605-672-37
18	159.5	10.4	1151	2	US-08-482-293A-37
19	159.5	10.4	1151	2	US-08-943-363-37
20	159.5	10.4	1151	4	US-09-193-043-37
21	159.5	10.4	1161	1	US-08-485-618-55
22	159.5	10.4	1161	1	US-08-362-652-55
23	159.5	10.4	1161	2	US-08-605-672-55
24	159.5	10.4	1161	2	US-08-482-293A-55
25	159.5	10.4	1161	2	US-08-943-363-55
26	159.5	10.4	1161	4	US-09-193-043-55
27	156	10.2	1161	1	US-08-173-497-2

28	156	10.2	1161	1	US-08-286-889-2	Sequence 2, Appl
29	156	10.2	1161	1	US-08-485-618-2	Sequence 2, Appl
30	156	10.2	1161	1	US-08-485-618-99	Sequence 99, Appl
31	156	10.2	1161	1	US-08-362-652-2	Sequence 2, Appl
32	156	10.2	1161	2	US-08-605-672-2	Sequence 2, Appl
33	156	10.2	1161	2	US-08-605-672-99	Sequence 99, Appl
34	156	10.2	1161	2	US-08-482-293A-2	Sequence 2, Appl
35	156	10.2	1161	2	US-08-482-293A-99	Sequence 99, Appl
36	156	10.2	1161	2	US-08-943-363-2	Sequence 2, Appl
37	156	10.2	1161	2	US-08-943-363-99	Sequence 99, Appl
38	156	10.2	1161	4	US-09-193-043-2	Sequence 2, Appl
39	156	10.2	1161	4	US-09-193-043-99	Sequence 99, Appl
40	149.5	9.7	435	5	PCT-US95-04439-1	Sequence 1, Appl
41	145.5	9.5	1152	2	US-08-476-062A-43	Sequence 43, Appl
42	145.5	9.5	1152	5	PCT-US96-01314-43	Sequence 43, Appl
43	145.5	9.5	1152	6	5424399-2	Patent No. 5424399
44	145.5	9.5	1153	1	US-08-173-497-3	Sequence 3, Appl
45	145.5	9.5	1153	1	US-08-286-889-3	Sequence 3, Appl

#### ALIGNMENTS

RESULT 1

US-08-286-889-46

: Sequence 46, Application US/08286889

: Patent No. 5470953

: GENERAL INFORMATION:

: APPLICANT: Gallatin, W. Mich

: APPLICANT: Van der Vieren, Monica

: TITLE OF INVENTION: NO. 5470953el Human 2 Integrin Alpha Subunit

: NUMBER OF SEQUENCES: 51

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

: STREET: 233 South Wacker Drive, 6300 Sear tower

: CITY: Chicago

: STATE: Illinois

: COUNTRY: United States

: ZIP: 60606-6402

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: Patent in Release #1.0, Version #1.25

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/286,889

: FILING DATE:

: CLASSIFICATION: 435

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: US 08/173,497

: FILING DATE: 23-DEC-1993

: ATTORNEY/AGENT INFORMATION:

: NAME: Williams Jr., Joseph A.

: REGISTRATION NUMBER: P38,659

: REFERENCE/DOCKET NUMBER: 27866/32168

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 312-474-6300

: TELEFAX: 312-474-0448

: TELEX: 25-3856

: INFORMATION FOR SEQ ID NO: 46:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 1155 amino acids

: TYPE: amino acid

: TOPOLOGY: linear

: MOLECULE TYPE: protein

: US-08-286-889-46

Query Match 10.8%; Score 166; DB 1; Length 1155;

Best Local Similarity 28.2%; Pred. No. 2.8e-09;

Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

Oy 11 PACYCG-GFDLYFILDKSGSV-LHHWNIYYFVEQLAHKFISPOLRMSFVIFSTRGTTLMK 68

Db 144 PECGQEMDIAFLIDSGSIDQSDFTQMKDFVKALMGQLASTSTSFSLMOYSNLIKTHFT 203  
Qy 69 LTEDR-----EQIRQGLEELQKVLPGDVTYMHGFERASEQIYYENRQGYRTA-SVITIAL 122  
Db 204 FTEFKSSLSQSLVDIAVOLQ-----GLTYTASGIQKVVKLEFHSKNGARKSAKKILIVI 258  
Qy 123 TDGELHEDLFFYSE--REANRSRDLGAIVYCVGKVD-FNE-TQLARI-----ADSKDHVF 173  
Db 259 TDGQKFRDPLEYRHVPEAKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSODHVF 315  
Qy 174 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVD 227  
Db 316 KVG-N-FVALRSIQRIQIEK-----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

RESULT 3

US-08-362-652-46  
; Sequence 46, Application US/08362652  
; Patent No. 5766850  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/362,652  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32391  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-362-652-46

Query Match 10.8%; Score 166; DB 1; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 2.8e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

Qy 11 PACYG-GFDLYFILDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMK 68  
Db 144 PECGQEMDIAFLIDSGSIDQSDFTQMKDFVKALMGQLASTSTSFSLMOYSNLIKTHFT 203  
Qy 69 LTEDR-----EQIRQGLEELQKVLPGDVTYMHGFERASEQIYYENRQGYRTA-SVITIAL 122

Db 144 PECGQEMDIAFLIDSGSIDQSDFTQMKDFVKALMGQLASTSTSFSLMOYSNLIKTHFT 203  
Qy 69 LTEDR-----EQIRQGLEELQKVLPGDVTYMHGFERASEQIYYENRQGYRTA-SVITIAL 122  
Db 204 FTEFKSSLSQSLVDIAVOLQ-----GLTYTASGIQKVVKLEFHSKNGARKSAKKILIVI 258  
Qy 123 TDGELHEDLFFYSE--REANRSRDLGAIVYCVGKVD-FNE-TQLARI-----ADSKDHVF 173  
Db 259 TDGQKFRDPLEYRHVPEAKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSODHVF 315  
Qy 174 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVD 227  
Db 316 KVG-N-FVALRSIQRIQIEK-----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

RESULT 2

US-08-485-618-46  
; Sequence 46, Application US/08485618  
; Patent No. 5728533  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485,618  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32797  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-485-618-46

Query Match 10.8%; Score 166; DB 1; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 2.8e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

Qy 11 PACYG-GFDLYFILDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMK 68

Db 204 FTEFKSSLPQSLVDAIVOLQ-----GLTYTASGIQKVVKELFHSKNGARKSAKKILIVI 258  
Qy 123 TDGELHEDLFYSE--REANRSRLGATVYCVGVKD-FNE-TQLARI-----ADSKDHVF 173  
Db 259 TDGQKFRDPLEYRHVIPAERKA---GIIRYAIGVGDAREPTALQELNTIGSAPSQDHVF 315  
Qy 174 PVNDFQALQGIHIIHLKSCIEILAAEPSTICAGESQVYVVRNGFRHARNVD 227  
Db 316 KVG-N-FVALRSIQRIQEK-----IFAIBGTESRSSSPQHEMSQEGFSSALSMD 364

RESULT 4  
US-08-605-672-46  
; Sequence 46, Application US/08605672  
; Patent No. 5817515  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: NO. 5817515el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/605,672  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32684  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-605-672-46

Query Match 10.8%; Score 166; DB 2; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 2.8e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;  
Qy 11 PACYGFGLDFILDKGSV-LHHWNEIYFVEQLAHKFISPOLRMSFIVFSTRTTLMK 68  
Db 144 PCPCQEMDIAFLIDSGSIDQSDFTQMKDFVKALMGQLASTSTFSLSMQYSNLIKTHFT 203  
Qy 69 LTEDR-----EQIRGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTA-SVIAL 122  
Db 204 FTEFKSSLPQSLVDAIVOLQ-----GLTYTASGIQKVVKELFHSKNGARKSAKKILIVI 258

Qy 123 TDGELHEDLFYSE--REANRSRLGATVYCVGVKD-FNE-TQLARI-----ADSKDHVF 173  
Db 259 TDGQKFRDPLEYRHVIPAERKA---GIIRYAIGVGDAREPTALQELNTIGSAPSQDHVF 315  
Qy 174 PVNDFQALQGIHIIHLKSCIEILAAEPSTICAGESQVYVVRNGFRHARNVD 227  
Db 316 KVG-N-FVALRSIQRIQEK-----IFAIBGTESRSSSPQHEMSQEGFSSALSMD 364  
RESULT 5  
US-08-482-293A-46  
; Sequence 46, Application US/08482293A  
; Patent No. 5831029  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: NO. 5831029el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/482,293A  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32684  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-482-293A-46

Query Match 10.8%; Score 166; DB 2; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 2.8e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;  
Qy 11 PACYGFGLDFILDKGSV-LHHWNEIYFVEQLAHKFISPOLRMSFIVFSTRTTLMK 68  
Db 144 PCPCQEMDIAFLIDSGSIDQSDFTQMKDFVKALMGQLASTSTFSLSMQYSNLIKTHFT 203  
Qy 69 LTEDR-----EQIRGLEELQKVLPGGDTYMHGFERASEQIYYENRQGYRTA-SVIAL 122  
Db 204 FTEFKSSLPQSLVDAIVOLQ-----GLTYTASGIQKVVKELFHSKNGARKSAKKILIVI 258

QY	123	TGDELHEDLPFYGE--REANRSDLGAIYVCVVKD-FNE-TQLARI-----ADSKDHVF	173
		:	
Db	259	TDGKFRDPLEYRHVIPEAEKA---GIIRYAIGVDAFREPTALQELNTIGSAPSDOHVF	315
QY	174	PVNDGFQALOGITHSLIKLKSCIEILAAEPSTICAGESFQVVVRCNGFRHARNVD	227
Db	316	KVGH-FVAIESTOTIOEK----TFATEGPSRSSSSFOHEMSOEFGFSALSMD	364

```

RESULT 6
US-08-943-363-46
; Sequence 46, Application US/08943363
; Patent No. 5837478
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,363
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1155 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-943-363-46

```

	Query Match	10.8%;	Score 166;	DB 2;	Length 1155;
	Best Local Similarity	28.2%;	Pred. No. 2.8e-09;		
	Matches	66;	Conservative 42;	Mismatches 96;	Indels 30; Gaps 12;
QY	11	PACYG-GFDLYFLDKSGSV-LHHWNEIYYFYEQLAHFISPOLRMSFIVSTGTTLMK	68		
Db	144	PECPGGEMDIAFLIDGSGIDQSDFTKMDFVKALMGOLASTSFSLMQRSNILKTHFT	203		
QY	69	LTEDR-----EQLRGLELKLQVPGGDYTMHEGFERASEQIYYENRCGYTA-SVIAL	122		
Db	204	FTEFKSSLSQSLDAIVOLQ-----GLTYTAGIQKVVELFHSHKGARKSAKKILIVI	258		
QY	123	TGDELHEDLFIFYSE--REANRRDLGAIIYCVGVKD-FNE-TQLARI-----ADSKDHVF	173		

```

Db      259 TDGQKFDPLBYRNVIPAEAKA---GIIRYAIGVDFAFREPTALQELNTIGSAPSQDHVF 315
        |||: | | | | | :|| :|| | | | | | :||| | | | | |
Qy      174 PVNDFOALOGIIHHSILKKSCIEILAAEPSTICAGESFVVVVRNGNTRFHARNVD 227
        |||: | | | | | :|| :|| | | | | | :|| | | | | |
Db      316 KVG-N-FVALRSIQIOEK----IFAETGETSRSSSFQHEMSOEGFSALSMD 364
        |||: | | | | | :|| :|| | | | | | :|| | | | | |

RESULT 7
US-09-193-043-46
; Sequence 46, Application US/09193043
; Patent No. 6251395
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: NO. 6251395el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/193,043
; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173,497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286,889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362,652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943,363
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: patentin ver. 2.0
; SEQ ID NO 46
; LENGTH: 1155
; TYPE: PRT
; ORGANISM: Mus musculus
JIS-09-193-043-46

```

	Query Match	10.8%;	Score 166;	DB 4:	Length 1155;	
	Best Local Similarity	28.2%;	Pred. No. 2.8e-09;			
	Matches	66;	Conservative 42;	Mismatches 96;	Indels 30;	Gaps 12;
QY	11	PACYG-GFDLYFILDKSGSV-LHHNWEIYFYVEQLAHKFIISPOLRMSFIVSTRTTLMK	68			
		: : : : : : : : : : : : : : : : : : : : : :				
Db	144	PCPCGOEMDIAFLIDGSSIGDSIQTKMFKALQMGLASTSTGSLMQVSNILKTFT	203			
		: : : : : : : : : : : : : : : : : : : : : :				
QY	69	LTEDR-----EQIROGLEELQVLPGDTVMHEGFPERASEQIYYENRGYRTA-SVITAL	122			
		: : : : : : : : : : : : : : : : : : : : : :				
Db	204	FTEFKSSLSPQLVAIVLQ-----GLTYTAGSQKVVKELFSKNKGARKSAKKILIVI	258			
		: : : : : : : : : : : : : : : : : : : : : :				
QY	123	TGDELHEDLFIFYSE--REANRSRDLCAIYCVGVKD-FNE-TQLARI-----ADSKDHVF	173			
		: : : : : : : : : : : : : : : : : : : : : :				
Db	259	TDGOKFRDPLEVRHVPIPEAKA---GIITYAIGVGDAFREPTALQELNTIGSAPSDHVF	315			
		: : : : : : : : : : : : : : : : : : : : : :				
QY	174	PVNDGFOALGGIIHTLKSCIEILAAEPPSTICAGESFOVVYRVNGFRHARNVD	227			
		: : : : : : : : : : : : : : : : : : : : : :				
Db	216	KYCN-EVALPSTOPOIFK----IFAETCESRSSSFQEHNSOGFESSALSMD	364			
		: : : : : : : : : : : : : : : : : : : : : :				

```

RESULT      8
US-08-485-618-53
; Sequence 53, Application US/08485618
; Patent No. 5728533
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5728533e1 Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
;

```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,618  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32797  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-485-618-53

Query Match 10.8%; Score 166; DB 1; Length 1161;  
Best Local Similarity 28.2%; Pred. No. 2.9e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

Qy 11 PACVG-GFDLYFLDKSGSV-LHHWNEYFYVEQLAHKFIQPLRMSTFVSTRTTLMK 68  
Db 144 PECQEQMDIAFLIDGSGSIDQSDFTQMKDFVKALMGQLASTSTFSLMQVSNILKTHFT 203  
Qy 69 LTEDR-----EQIROGLELQKLVPGGDTYMHGEPERASEQIYYENRQGYRTA-SVIAL 122  
Db 204 FTEFKSSLPQSLVDIAVLQ-----GLTYTASGIQKVVKELFHSKNGARKSAKKILIVI 258  
Qy 123 TDGELHEDLFFYSE--REANRSRLGAIYVCVGVKD-FNE-TQLARI-----ADSKDHVF 173  
Db 259 TDGQKFRDPLEYRHYVPEAKA--GIIRYAIGVGDAFREPTALQELNTIGSAPSQDHVF 315  
Qy 174 PVNDGFOALQGLIHSILKKSCIEILAAEPSTICAGESFQVVVVRNGFRHARNVD 227  
Db 316 KVG-N-FVALRSIQRIQIEK----IPAIEGTSSRSSSFQHEMSQEGFSSALSMD 364

RESULT 9  
US-08-362-652-53  
Sequence 53, Application US/08362652  
Patent No. 5766850  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 93  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/362,652  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32391  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-362-652-53

Query Match 10.8%; Score 166; DB 1; Length 1161;  
Best Local Similarity 28.2%; Pred. No. 2.9e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

Qy 11 PACVG-GFDLYFLDKSGSV-LHHWNEYFYVEQLAHKFIQPLRMSTFVSTRTTLMK 68  
Db 144 PECQEQMDIAFLIDGSGSIDQSDFTQMKDFVKALMGQLASTSTFSLMQVSNILKTHFT 203  
Qy 69 LTEDR-----EQIROGLELQKLVPGGDTYMHGEPERASEQIYYENRQGYRTA-SVIAL 122  
Db 204 FTEFKSSLPQSLVDIAVLQ-----GLTYTASGIQKVVKELFHSKNGARKSAKKILIVI 258  
Qy 123 TDGELHEDLFFYSE--REANRSRLGAIYVCVGVKD-FNE-TQLARI-----ADSKDHVF 173  
Db 259 TDGQKFRDPLEYRHYVPEAKA--GIIRYAIGVGDAFREPTALQELNTIGSAPSQDHVF 315  
Qy 174 PVNDGFOALQGLIHSILKKSCIEILAAEPSTICAGESFQVVVVRNGFRHARNVD 227  
Db 316 KVG-N-FVALRSIQRIQIEK----IPAIEGTSSRSSSFQHEMSQEGFSSALSMD 364

RESULT 10  
US-08-605-672-53  
Sequence 53, Application US/08605672  
Patent No. 5817515  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

APPLICATION NUMBER: US/08/482,293A  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-482-293A-53

Query Match 10.8%; Score 166; DB 2; Length 1161;  
Best Local Similarity 28.2%; Pred. No. 2.9e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 11 PACYG-GFDLYFLDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFIVFSTGTTLMK 68  
DB 144 PCPCQEMDIAFLIDGSGSIDQSDFTQMKDFVKALMGQLASTSTFSLMOYSNLIKTHFT 203  
QY 69 LTEDR-----EQIRGLELEQLKVLPGDVTYMHGFERASEQIYYENROGYRTA-SVIAL 122  
DB 204 FTEFKSSLSQSLDAIVQLQ-----GLITYASGIQKVVKELFHSKNGARKSAKKILIVI 258  
QY 123 TDGELHEDLFFYSE--REANRSRLGALVYCVGVKD-FNE-TQLARI-----ADSKDHVF 173  
DB 259 TDGQKRPDLFHRVPIPEAKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSQDHVF 315  
QY 174 PVNDGFALQGIHSLTKKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVD 227  
DB 316 KVGNFVALRSIQRIQEK----IFAIEGTSRSSSFQHEMSQEGFSSALSMD 364

RESULT 12  
US-08-943-363-53  
Sequence 53, Application US/08943363  
Patent No. 5837478  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 114  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/943.363

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/605,672  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-605-672-53

Query Match 10.8%; Score 166; DB 2; Length 1161;  
Best Local Similarity 28.2%; Pred. No. 2.9e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 11 PACYG-GFDLYFLDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFIVFSTGTTLMK 68  
DB 144 PCPCQEMDIAFLIDGSGSIDQSDFTQMKDFVKALMGQLASTSTFSLMOYSNLIKTHFT 203  
QY 69 LTEDR-----EQIRGLELEQLKVLPGDVTYMHGFERASEQIYYENROGYRTA-SVIAL 122  
DB 204 FTEFKSSLSQSLDAIVQLQ-----GLITYASGIQKVVKELFHSKNGARKSAKKILIVI 258  
QY 123 TDGELHEDLFFYSE--REANRSRLGALVYCVGVKD-FNE-TQLARI-----ADSKDHVF 173  
DB 259 TDGQKRPDLFHRVPIPEAKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSQDHVF 315  
QY 174 PVNDGFALQGIHSLTKKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVD 227  
DB 316 KVGNFVALRSIQRIQEK----IFAIEGTSRSSSFQHEMSQEGFSSALSMD 364

RESULT 11  
US-08-482-293A-53  
Sequence 53, Application US/08482293A  
Patent No. 5831029  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

```
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-943-363-53

Query Match 10.8%; Score 166; DB 2; Length 1161;
Best Local Similarity 28.2%; Pred. No. 2.9e-09;
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

Qy 11 PACYG-GFDLYFILDKSGSV-LHHWNEIYFYVEQLAHKFISPOLRMSFIVFSTGTTLMK 68
Db 144 PECQEMDIAFLDGSIDQSDFTQKDFVKALMGQLASTSFSLMQVSNILKTHFT 203
Qy 69 LTEDR-----EQIROGLEELQKVLPGDPTMHGCFERASEQIYYENRQGYRTA-SVIAL 122
Db 204 FTEFKSSLSQSLVDIAVOLQ-----GLTYTASGIQKVVKELFHSKNGARKSAKKILIVI 258
Qy 123 TDGELHEDLFFYSE--REANRSRDLGAIVYCVGVKD-FNE-TQLARI-----ADSKDHVF 173
Db 259 TDGQKFRDPLEYRHRVPEAKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSQDHVF 315
Qy 174 PVNDGFOALQGIHLSILKSCIEILAAEPSTICAGESFQVVVVRGNGFRHARNVD 227
Db 316 KVG-N-FVALRSIQRIQEK----IPAEIGTSRSSSSSFQHEMSQEGFSSALSMD 364

RESULT 13
US-09-193-043-53
; Sequence 53, Application US/09193043
; Patent No. 6251395
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 6251395el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/193,043
; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173,497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286,889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362,652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943,363
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
; LENGTH: 1161
; TYPE: PRT
```

```
; ORGANISM: Mus musculus
US-09-193-043-53

Query Match 10.8%; Score 166; DB 4; Length 1161;
Best Local Similarity 28.2%; Pred. No. 2.9e-09;
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

Qy 11 PACYG-GFDLYFILDKSGSV-LHHWNEIYFYVEQLAHKFISPOLRMSFIVFSTGTTLMK 68
Db 144 PECQEMDIAFLDGSIDQSDFTQKDFVKALMGQLASTSFSLMQVSNILKTHFT 203
Qy 69 LTEDR-----EQIROGLEELQKVLPGDPTMHGCFERASEQIYYENRQGYRTA-SVIAL 122
Db 204 FTEFKSSLSQSLVDIAVOLQ-----GLTYTASGIQKVVKELFHSKNGARKSAKKILIVI 258
Qy 123 TDGELHEDLFFYSE--REANRSRDLGAIVYCVGVKD-FNE-TQLARI-----ADSKDHVF 173
Db 259 TDGQKFRDPLEYRHRVPEAKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSQDHVF 315
Qy 174 PVNDGFOALQGIHLSILKSCIEILAAEPSTICAGESFQVVVVRGNGFRHARNVD 227
Db 316 KVG-N-FVALRSIQRIQEK----IPAEIGTSRSSSSSFQHEMSQEGFSSALSMD 364

RESULT 14
US-08-286-889-37
; Sequence 37, Application US/08286889
; Patent No. 5470953
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Mich
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/286,889
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: P38,659
; REFERENCE/DOCKET NUMBER: 27866/32168
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-286-889-37

Query Match 10.4%; Score 159.5; DB 1; Length 1151;
Best Local Similarity 27.9%; Pred. No. 1.5e-08;
Matches 63; Conservative 41; Mismatches 93; Indels 29; Gaps 11;
```

QY 18 DLYFTLDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFIVFSTGRGTTLMKLTE----- 71  
Db 142 DIAFLIDGSGSINQORDFAQMKDFVKALMGFEFASTLFSLMQYSNLIKTHFTTFEKNIL 201  
QY 72 DREQIROGLEELQKVLPGDGYMHGFEFASQIYYENRQGYRTA-SVIALTDGELHED 130  
Db 202 DPQSLVDPIVQLQ-----GLTYTATGIRTVMEELFHSKNGSRKSAKKILLVITDGQKYRD 256  
QY 131 LFFYSE--REANRSRDLGAIVCVGKVD-FNE-TOLARI-----ADSKDHVPVNDGFOA 181  
Db 257 PLEYSVDVIPAADKA---GIIRYAIGVGDAFOEPTALKELNTIGSAPPQDHFVKVGN-FAA 312  
QY 182 LOGIIHSLKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVD 227  
Db 313 LRSIQRLQLEK----IFAIEGTQSRSSSFQHEMSQEGFSSALTS 354

Search completed: August 9, 2002, 10:33:05  
Job time: 184 sec

QY 18 DLYFTLDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFIVFSTGRGTTLMKLTE----- 71  
Db 142 DIAFLIDGSGSINQORDFAQMKDFVKALMGFEFASTLFSLMQYSNLIKTHFTTFEKNIL 201  
QY 72 DREQIROGLEELQKVLPGDGYMHGFEFASQIYYENRQGYRTA-SVIALTDGELHED 130  
Db 202 DPQSLVDPIVQLQ-----GLTYTATGIRTVMEELFHSKNGSRKSAKKILLVITDGQKYRD 256  
QY 131 LFFYSE--REANRSRDLGAIVCVGKVD-FNE-TOLARI-----ADSKDHVPVNDGFOA 181  
Db 257 PLEYSVDVIPAADKA---GIIRYAIGVGDAFOEPTALKELNTIGSAPPQDHFVKVGN-FAA 312  
QY 182 LOGIIHSLKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVD 227  
Db 313 LRSIQRLQLEK----IFAIEGTQSRSSSFQHEMSQEGFSSALTS 354

RESULT 15  
US-08-485-618-37  
; Sequence 37, Application US/08485618  
; Patent No. 5728533  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: NO. 5728533el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485,618  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32797  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 37:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1151 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-485-618-37

Query Match 10.4%; Score 159.5; DB 1; Length 1151;  
Best Local Similarity 27.9%; Pred. No. 1.5e-08;  
Matches 63; Conservative 41; Mismatches 93; Indels 29; Gaps 11;



---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:34:12 ; Search time 61.73 Seconds  
(without alignments)  
459.199 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_27\_321

Perfect score: 1536

Sequence: 1 GQGRREDGGPACVGGFDLY.....GLSFSSSVIIITTHCSDGS 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Watch 0%

Maximum Watch 100%

Listing first 45 summaries

Database :

PIR\_71:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results, predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	145.5	9.5	1153	1 RWHU1B	cell surface glyco
2	143.5	9.3	1163	1 RWHU1C	cell surface glyco
3	143	9.3	3124	1 A40020	collagen alpha 1(X
4	141	9.2	1153	2 S00551	leukocyte surface
5	139	9.0	724	2 A48569	antigen Em100 - Ei
6	137	8.9	1747	1 A45974	collagen alpha 1(X
7	134.5	8.8	712	2 A45638	immunodominant mic
8	131	8.5	760	1 C2MS	classical-compleme
9	130	8.5	1857	2 S31212	collagen alpha 1(X
10	130	8.5	1888	2 S78476	collagen alpha 1(X
11	128	8.3	3051	2 S42373	hypothetical prote
12	118	7.7	929	2 I51027	type XII collagen
13	113.5	7.4	456	2 A37979	cartilage matrix p
14	113.5	7.4	704	1 BBHU	complement factor
15	111.5	7.3	2944	2 A54849	collagen alpha 1(V
16	110.5	7.2	761	1 BBMS	complement factor
17	109.5	7.1	1029	1 S21369	collagen alpha 2(V
18	108.5	7.1	500	2 S66522	cartilage matrix p
19	107.5	7.0	493	2 A33809	cartilage matrix p
20	105.5	6.9	567	2 T28797	hypothetical prote
21	104	6.8	747	2 I51579	complement factor
22	100.5	6.5	574	2 A46283	sporozoite surface
23	100	6.5	460	2 T23087	hypothetical prote
24	99.5	6.5	597	2 S33578	rop protein - frui
25	99	6.4	559	2 S04531	thrombospondin-rel
26	99	6.4	698	2 B96958	dnak protein (heat
27	99	6.4	817	2 T50240	kinesin-like prote
28	99	6.4	917	2 S09646	collagen alpha 2(V
29	99	6.4	1018	1 CGHU2A	collagen alpha 2(V

30	98.5	6.4	537	2 T04822	hypothetical prote
31	98	6.4	741	2 T46488	hypothetical prote
32	98	6.4	3176	1 CGHU3A	collagen alpha 3(V
33	97.5	6.3	642	2 H81185	dnak protein NMB05
34	97	6.3	3137	1 A37797	collagen alpha 3(V
35	96.5	6.3	676	2 T47637	hypothetical prote
36	96	6.2	272	2 A55348	integrin alpha-1 -
37	96	6.2	689	2 F84811	probable retroelem
38	95.5	6.2	334	2 AF1166	transcription regu
39	95.5	6.2	932	2 JCS953	inter-alpha-inhibi
40	95.5	6.2	1179	2 A53213	integrin alpha-E c
41	94.5	6.2	1151	2 A45226	integrin alpha-1 c
42	94	6.1	843	2 A40970	undulin 1 - human
43	93	6.1	340	2 E70121	hypothetical prote
44	92.5	6.0	642	2 B81917	probable chaperone
45	92	6.0	292	2 B83736	transposase (04) B

#### ALIGNMENTS

##### RESULT 1

RWHU1B

cell surface glycoprotein CD11b precursor [validated] - human

N:Alternate names: complement receptor type 3 alpha chain; Leukocyte adhesion protein  
eukocyte integrin alpha chain; neutrophil adherence receptor alphas chain

C:Species: Homo sapiens (man)

C:date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 31-Dec-2000

R:Corbi, A.L.; Kishimoto, T.K.; Miller, L.J.; Springer, T.A.

J. Biol. Chem. 263, 12403-12411, 1988

A:title: The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3,  
B.

A:Reference number: A31108; MUID:88315033

A:Accession: A31108

A:Molecule type: mRNA

A:Residues: 1-1153 <CR>

A:Cross-references: GB:J03925; NID:g187284; PIDN:AAA59544.1; PID:g307148

A>Note: part of this sequence was confirmed by protein sequencing

R:Arnaout, M.A.; Gupta, S.K.; Pierce, M.W.; Tenen, D.G.

J. Cell Biol. 106, 2153-2158, 1988

A:title: Amino acid sequence of the alpha subunit of human leukocyte adhesion recepto

A:Reference number: A28915; MUID:88257215

A:Accession: A28915

A:Molecule type: mRNA

A:Residues: 1-499,501-965,'P',967-1153 <ARN>

A:Cross-references: GB:M18044; GB:J03270; GB:M19664; GB:X07421; NID:g186935; PIDN:AAA

A>Note: the authors translated the codon TAC for residue 1129 as Thr

A>Note: part of this sequence, including the amino end of the mature protein, was con

R:Shelley, C.S.; Arnaout, M.A.

Proc. Natl. Acad. Sci. U.S.A. 88, 10525-10529, 1991

A:title: The promoter of the CD11b gene directs myeloid-specific and developmentally

A:Reference number: A41600; MUID:92073318

A:Accession: A41600

A:Molecule type: DNA

A:Residues: 1-9 <SHE>

A:Cross-references: GB:M76724; NID:g180018; PIDN:AAA58410.1; PID:g553215

R:Arnaout, M.A.; Remold-O'Donnell, E.; Pierce, M.W.; Harris, P.; Tenen, D.G.

Proc. Natl. Acad. Sci. U.S.A. 85, 2776-2780, 1988

A:title: Molecular cloning of the alpha-subunit of human and guinea pig leukocyte adh

A:Reference number: A94193; MUID:88190151

A:Accession: A30892

A:Molecule type: mRNA

A:Residues: 917-1042 <AR2>

A:Cross-references: GB:M18044

R:Hickstein, D.D.; Hickey, M.J.; Ozols, J.; Baker, D.M.; Back, A.L.; Roth, G.J.

Proc. Natl. Acad. Sci. U.S.A. 86, 257-261, 1989

A:title: cDNA sequence for the alphas subunit of the human neutrophil adherence recep

A:Reference number: A32218; MUID:89098893

A:Accession: A32218

A:Molecule type: mRNA

A:Residues: 9-1153 <HIC>

A:Cross-references: GB:J04145; NID:g189068; PIDN:AAA59903.1; PID:g386975

cell surface glycoprotein CD11c precursor - human  
 N:Alternate names: leukocyte adhesion receptor p150,95 alpha chain  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 22-Jun-1999  
 C:Accession: A35584; A35543; S00864  
 R:Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.  
 J. Biol. Chem. 265, 12750-12751, 1990  
 A:Reference number: A36584  
 A:Contents: erratum  
 A:Accession: A36584  
 A:Molecule type: DNA  
 A:Residues: 1-1163 <COR>  
 A:Note: this revision to the sequence from reference A35543 includes the carboxyl end  
 R:Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.  
 J. Biol. Chem. 265, 2782-2788, 1990  
 A:Title: Genomic structure of an integrin alpha subunit, the leukocyte p150,95 molecu  
 A:Reference number: A35543; MUID:90153906  
 A:Accession: A35543  
 A:Molecule type: DNA  
 A:Residues: 1-834 <CO2>  
 A:Note: this sequence has been revised in reference A36584  
 R:Corbi, A.L.; Miller, L.J.; O'Connor, K.; Larson, R.S.; Springer, T.A.  
 EMBO J. 6, 4023-4028, 1987  
 A:Title: cDNA cloning and complete primary structure of the alpha subunit of a leukoc  
 A:Reference number: S00864; MUID:88166645  
 A:Accession: S00864  
 A:Molecule type: mRNA  
 A:Residues: 1-755, 'L', 757-1163 <CO3>  
 A:Cross-references: GB:M81695; EMBL:Y00093; NID:g487829; PIDN:AAA59180.1; PID:g487830  
 A:Note: part of this sequence was confirmed by protein sequencing  
 C:Comment: A common beta chain (CD18) forms a heterodimer with CD11b to form Mac-1 on  
 C:Genetics:  
 A:Gene: GDB:ITGAX; CD11C  
 A:Cross-references: GDB:119758; OMIM:151510  
 A:Map position: 16p11.2-16p11.2  
 C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat h  
 C:Keywords: calcium; cell adhesion; glycoprotein; heterodimer; magnesium; tandem repe  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-1163/Product: cell surface glycoprotein CD11c #status predicted <MAT>  
 F:20-1163/Product: cell surface glycoprotein CD11c #status predicted <EXT>  
 F:20-1163/Product: cell surface glycoprotein CD11c #status predicted <WMA4>  
 F:149-319/Domain: von Willebrand factor type A repeat homology <WMA4>  
 F:1108-1133/Domain: extracellular #status predicted <EXT>  
 F:1108-1133/Domain: transmembrane #status predicted <WMA4>  
 F:1134-1163/Domain: intracellular #status predicted <INT>  
 F:61,89,392,697,735,899,939,1050/Binding site: carbohydrate (Asn) (covalent) #status

Query Match 9.5%; Score 145.5; DB 1; Length 1153;  
 Best Local Similarity 26.2%; Pred. No. 0.0014;  
 Matches 60; Conservative 45; Mismatches 81; Indels 43; Gaps 12;  
 QY 18 DLYFILDKSGSV-LHHNNEIYFVEQLAHKFIISPOLRMSFIVFSTRGTTLMKLTEDRE-- 74  
 Db 151 DIVFLIDGSGSISSRNFAFMNFRVAVISQFQRPSTQFSLMOFSNKFQTHFTFEFRRTS 210  
 QY 75 ---QIROGLEELQKVLPGDGYMHGFEFRASEIOIYVENRQGYRTAS-VIIATDGLHED 130  
 Db 211 NPLSLASVHQLQ-----GFTYATATQNVVHRLFHASYGARDATKILIVTDGKKEG 265  
 QY 131 LFFYSERANRSDLAGIAIVYCVGV-----KDFNETQLARIAD--SKDHVFPVNDG 178  
 Db 266 SLDYKD-VIPMADAAGIIRYAIGVGLAFQNRNSWKLND-----IAKPSQEHFKVED- 318  
 QY 179 FOALOGIHSILKSCIEILAAEPSTICAGSFQVVRNGNF 220  
 Db 319 FDALKDIONQLKEK-----IFAEGTETTSSTSSFELEMAQEGF 356

RESULT 3  
 A40020  
 collagen alpha 1(XII) chain precursor - chicken  
 N:Alternate names: fibronectin  
 C:Species: Gallus gallus (chicken)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 19-Jan-2001

A:Note: part of this sequence was confirmed by protein sequencing  
 R:Fleming, J.C.; Pahl, H.L.; Gonzalez, D.A.; Smith, T.F.; Tenen, D.G.  
 J. Immunol. 150, 480-490, 1993  
 A:Title: Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-in  
 n during evolution.  
 A:Reference number: A46526; MUID:93123748  
 A:Accession: A46526  
 A:Status: not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-499,501-1153 <FLE>  
 A:Cross-references: GB:S2227; NID:g263047; PIDN:AA24821.1; PID:g263049  
 A:Note: the last three bases of intron 13, CAG, are included in some but not all mature  
 A:Note: sequence extracted from NCBI backbone (NCBIP:121963)  
 R:Pierce, M.W.; Remold-O'Donnell, E.; Todd III, R.F.; Arnout, M.A.  
 Blochim. Biophys. Acta 874, 368-371, 1986  
 A:Title: N-terminal sequence of human leukocyte glycoprotein Mol: conservation across sp  
 A:Reference number: A90664; MUID:87076671  
 A:Accession: A26091  
 A:Molecule type: protein  
 A:Residues: 17-31 <PIE>  
 A:Experimental source: granulocytes  
 R:Pahl, H.L.; Rosmarin, A.G.; Tenen, D.G.  
 Blood 79, 865-870, 1992  
 A:Title: Characterization of the myeloid-specific CD11b promoter.  
 A:Reference number: 152567; MUID:92144986  
 A:Accession: 152567  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-9 <RES>  
 A:Cross-references: GB:M84477; NID:g180184; PIDN:AA51960.1; PID:g553219  
 C:Comment: A common beta chain (CD18) forms a heterodimer with this chain to form Mac-1  
 C:Genetics:  
 A:Gene: GDB:ITGAM; CR3A  
 A:Cross-references: GDB:120599; OMIM:120980  
 A:Map position: 16p11.2-16p11.2  
 A:Note: promoter contains a GATA motif and two Spl consensus binding sites  
 C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat hom  
 C:Keywords: alternative splicing; calcium; cell adhesion; glycoprotein; heterodimer; mag  
 F:1-16/Domain: signal sequence #status predicted <SIG>  
 F:17-1153/Product: cell surface glycoprotein CD11b #status experimental <MAT>  
 F:17-1153/Product: extracellular #status predicted <EXT>  
 F:148-318/Domain: von Willebrand factor type A repeat homology <WMA2>  
 F:465-473/Region: calcium/magnesium binding #status predicted  
 F:530-538/Region: calcium/magnesium binding #status predicted  
 F:593-601/Region: calcium/magnesium binding #status predicted  
 F:1109-1134/Domain: transmembrane #status predicted <WMA4>  
 F:1135-1153/Domain: intracellular #status predicted <INT>  
 F:86,240,391,469,693,697,735,802,881,901,912,941,947,979,994,1022,1045,1051,1076/Binding



Db	102	SDSRAQNADLLAAAKKLPYAAAGSTYTHLGLAKA--EELIFSFGKGRDNAPKMLVMTDG	160
QY	126	ELHEDLFFYSERE-----ANRSDLGAIYVGVG-KDFNETQLARIA--DSKDHV-FP-	174
Db	161	A-----SSRSQTLSSAAEKLNRGVIIVLVGCTGVNSAECSIAIGCDTSDTVECPR	212
QY	175	-VNDGFOALOGIIHSLKKSCEIEI-----LAAPSTI--CAGE	209
Db	213	YLOSNWGVSSQINGIIKACKDLAKDAVCSEWSEYGPCEGE	254
RESULT 6			
A45974			
collagen alpha 1(XIV) chain precursor, short form 2 - chicken			
N:Alternate names: undulin			
C:Species: Gallus gallus (chicken)			
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999			
C:Accession: A45974; S30085; S22916; S17035; S20833			
R:Gerecke, D.R.; Foley, J.W.; Castagnola, P.; Gennari, M.; Dublet, B.; Cancedda, R.;			
J. Biol. Chem. 268, 12177-12184, 1993			
A:Title: Type XIV collagen is encoded by alternative transcripts with distinct 5' reg			
ns.			
A:Reference number: A45974; MUID:93280195			
A:Accession: A45974			
A:Status: preliminary			
A:Molecule type: mRNA; protein			
A:Residues: 1-1747 <GER>			
A:Experimental source: embryo skin			
A:Note: sequence inconsistent with the nucleotide translation			
A:Note: sequence extracted from NCBI backbone (NCBIN:133364, NCBIPI:133365)			
R:Apte, S.S.			
submitted to the EMBL Data Library, March 1992			
A:Reference number: S30085			
A:Accession: S30085			
A:Molecule type: mRNA			
A:Residues: 1472-1660 <APT>			
A:Cross-references: EMBL:X65122; NID:g62871; PIDN:CAA46238.1; PID:g938175			
R:Trueb, J.; Trueb, B.			
Eur. J. Biochem. 207, 549-557, 1992			
A:Title: Type XIV collagen is a variant of undulin.			
A:Reference number: S22916; MUID:92339443			
A:Accession: S22916			
A:Status: preliminary			
A:Molecule type: mRNA			
A:Residues: 286-494, 'Q', 496-834, 'A', 836-1119, 'KL', 1122-1402, 1409-1439 <TRU>			
R:Gordon, M.K.; Castagnola, P.; Dublet, B.; Linsenmayer, T.F.; van der Rest, M.; Mayn			
Eur. J. Biochem. 201, 333-338, 1991			
A:Title: Cloning of a cDNA for a new member of the class of fibril-associated collage			
A:Reference number: S17035; MUID:92037585			
A:Accession: S17035			
A:Molecule type: mRNA			
A:Residues: 1472-1659 <GOR1>			
A:Accession: S20833			
A:Molecule type: protein			
A:Residues: 1551-1570; 1593-1599; 1639-1667 <GOR2>			
C:Superfamily: collagen alpha 1(XIV) chain; fibronectin type III repeat homology; von			
C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; tr			
F:40-204/Domain: von Willebrand factor type A repeat homology <FN3A>			
F:236-317/Domain: fibronectin type III repeat homology <FN3B>			
F:326-409/Domain: fibronectin type III repeat homology <FN3C>			
F:418-498/Domain: fibronectin type III repeat homology <FN3D>			
F:507-591/Domain: fibronectin type III repeat homology <FN3E>			
F:625-707/Domain: fibronectin type III repeat homology <FN3F>			
F:716-798/Domain: fibronectin type III repeat homology <FN3G>			
F:806-893/Domain: fibronectin type III repeat homology <FN3H>			
F:924-1089/Domain: von Willebrand factor type A repeat homology <VWA2>			
F:1111-1352/Domain: non-collagenous NC4 #status predicted <NC4>			
F:1511-1553/Domain: non-collagenous NC2 #status predicted <NC2>			
F:1554-1659/Domain: triple helical domain COL1 #status predicted <COL1>			
Query Match 8.9%; Score 137; DB 1; Length 1747;			
Best Local Similarity 25.4%; Pred. No. 0.012;			

A:Residues: 11-44 <RES>		
A:Cross-references: GB:M14293; NID:gl98993; PIDN:AAA39484.1; PID:g554193		
C:Gene: Mac-1		
C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo		
C:Keywords: cell adhesion; glycoprotein; transmembrane protein		
F:1-16/Domain: signal sequence #status predicted <SIG>		
F:17-1153/Product: leukocyte surface glycoprotein Mac-1 alpha chain #status experimental		
F:148-318/Domain: von Willebrand factor type A repeat homology <VWA2>		
F:1106-1129/Domain: transmembrane #status predicted <TM>		
Query Match 9.2%; Score 141; DB 2; Length 1153;		
Best Local Similarity 24.2%; Pred. No. 0.0033;		
Matches 71; Conservative 50; Mismatches 96; Indels 76; Gaps 16;		
QY	18 DLYFTLDKSGVLHWNIIYFVEQLAHKFTS---PQLRMSFIYFSTGRTITLMLKLTED--	72
Db	150 DIVFLIDGSGSI---NNIDF---QMKFEVSVIMEQFKSKTLFTS-----LMQYSDEPR	197
QY	73 ---REQIROGLELEQKVLTP---GGDTYMHGEGFERASEQIYFE-NRQGYRTASVITALT	123
Db	198 IHFTFNDEKFRNPSPSHVSPKQLNGRTKTASGIRKVVVRELFHKTNGARENAKILVVIT	257
QY	124 DGELHEDLFFYSE--REANRSDLGAIYVGVGD-FNETQLARIAD-----SKDHVEP	174
Db	258 DGEKFGDPLDYKQVPEADRA---GVIRYVIGVGNFKNPQSRRELDTIASKPAGEHFVQ	314
QY	175 VNDGFOALOGIHTSLKKSCEIELAAEPSTICAGESFQVVRGNGFHRARNVDRVRLCSFK	234
Db	315 V-DNFELANTIONQLEK---IFAIEGTQGTSTSFSEHMSQEGF-----	355
QY	235 INDSTVLTNEKPFVSVDYLLCPAPILKEVG---MKAALQVSMNDGLSFISSS	283
Db	356 -SASITSN-----GPLLGSVGSFDWAGGAFLYTSKDKVTFINTT	393
RESULT 5		
A48569		
antigen Em100 - Eimeria maxima		
C:Species: Eimeria maxima		
C:Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998		
C:Accession: A48569		
R:Pasamonies, L.; Hug, D.; Humbelin, M.; Weber, G.		
Mol. Biochem. Parasitol. 57, 171-174, 1993		
A:Title: Sequence of a major Eimeria maxima antigen homologous to the Eimeria tenella m		
A:Reference number: A48569; MUID:93149203		
A:Accession: A48569		
A:Status: preliminary		
A:Molecule type: nucleic acid		
A:Residues: 1-724 <PAS>		
A:Cross-references: GB:M99058; NID:gl58890; PID:gl58891		
C:Superfamily: thrombospondin type 1 repeat homology; von Willebrand factor type A repea		
F:45-218/Domain: von Willebrand factor type A repeat homology <VWA2>		
F:238-296/Domain: thrombospondin type 1 repeat homology <THR1>		
F:309-371/Domain: thrombospondin type 1 repeat homology <THR2>		
F:372-432/Domain: thrombospondin type 1 repeat homology <THR3>		
F:433-493/Domain: thrombospondin type 1 repeat homology <THR4>		
F:494-556/Domain: thrombospondin type 1 repeat homology <THR5>		
F:560-610/Domain: thrombospondin type 1 repeat homology <THR6>		
Query Match 9.0%; Score 139; DB 2; Length 724;		
Best Local Similarity 26.6%; Pred. No. 0.0026;		
Matches 59; Conservative 40; Mismatches 89; Indels 34; Gaps 14;		
QY	13 CYGGFDLYFILDKSGSV-LHWNIIYFVEQLAHKF-ISP-QLRMSFIYFSTGRTITLMLK	69
Db	42 CTRLLDVMVLVDESGISGTSYNGYFRFISNFACTMPLSPDDVRVGLVTFGTSAVTRWDL	101
QY	70 TEDRQIROGLELEQKVLTP---GGDTYMHGEGFERASEQIYENRQGYR--TASVITALT	125



```

QY 273 MND-----GLSPFISSVII 286
      : | | | | |
Db 531 VGDPTSQHGKELVEDVII 549

RESULT 9
S31212
collagen alpha 1(XIV) chain precursor, short form - chicken
C:Species: Gallus gallus (chicken)
C:Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 13-Aug-1999
C:Accession: S31212
R:Waeichli, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.
Eur. J. Biochem. 212, 483-490, 1993
A:Title: Complete primary structure of chicken collagen XIV.
A:Reference number: S31211; MUID:93185668
A:Accession: S31212
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-1857 <WAE>
A:Cross-references: EMBL:X70792; NID:g288874; PIDN:CAA50063.1; PID:g288875
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C:Genetics:
A:Gene: Coll14A1
C:Superfamily: collagen alpha 1(XIV) chain; fibronectin type III repeat homology; von Willebrand factor type A repeat homology
C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; triple helix
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-110/Domain: fibronectin type III repeat homology <FN3A>
F:156-320/Domain: von Willebrand factor type A repeat homology <VWA1>
F:442-525/Domain: fibronectin type III repeat homology <FN3C>
F:534-614/Domain: fibronectin type III repeat homology <FN3D>
F:623-707/Domain: fibronectin type III repeat homology <FN3E>
F:741-823/Domain: fibronectin type III repeat homology <FN3F>
F:832-914/Domain: fibronectin type III repeat homology <FN3G>
F:922-1009/Domain: fibronectin type III repeat homology <FN3H>
F:1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 8.5%; Score 130; DB 2; Length 1857;
Best Local Similarity 25.0%; Pred. No. 0.049; Indels 26; Gaps 12;
Matches 62; Conservative 44; Mismatches 116;

QY 18 DLYFTLDKSGSV-LHWNHNIYFVEQL--AHKFISPO-LRMSFIVFSTGTMLMKT--E 71
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db 1042 DLVFLVDGWSIGDDNFNKIISFLYSTVGALDKIGPDGTQVAIIQFSDDPRTFEKLNAYK 1101

QY 72 DREIQRGLEELQKVLPGDGYMHGFEFRASEQIYENRQGYRTA--SVIIALTGDGELHE 129
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
Db 1102 TKETLLEAIQOI--AYKGGNTKTKRAIKHARE-VLFTGAGMRKGIPIKVLVITDGRSQD 1158

QY 130 DLFFYSEREANRDL-GAIVYCVGVKDFNETOLARIAD--SKDHVFPVNDGFQALOGII 186
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
Db 1159 DV-----NKVSRMQLDGFSAIGVADADYSELVNIQSKPSERHVFVDD-DFDAFTKIE 1212

QY 187 HSILKKSCEILAAEPSTICAGESFQVVVGRNGFRHARNVDRVLCFSKINDSVTLNEKPF 246
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
Db 1213 DELITFVCETASATCPLVFKDGDKLA-----GFKMMEMFGLVEKEFSALDGVSMPEPGTF 1266

QY 247 SVEDTYLL 254
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
Db 1267 NVYPCYRL 1274

RESULT 10
S78476
collagen alpha 1(XIV) chain precursor, long form - chicken
C:Species: Gallus gallus (chicken)
C:Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 13-Aug-1999
C:Accession: S78476; S31211
R:Trueb, B.
submitted to the EMBL Data Library, January 1993

A:Reference number: S78476
A:Accession: S78476
A:Molecule type: mRNA
A:Residues: 1-1888 <TRU>
A:Cross-references: EMBL:X70793; NID:g288872; PIDN:CAA50064.1; PID:g288873
R:Waeichli, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.
Eur. J. Biochem. 212, 483-490, 1993
A:Title: Complete primary structure of chicken collagen XIV.
A:Reference number: S31211; MUID:93185668
A:Accession: S31211
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-416;1460-1811,1843-1888 <WAE>
A:Cross-references: EMBL:X70793
C:Genetics:
A:Gene: Coll14A1
C:Superfamily: collagen alpha 1(XIV) chain; fibronectin type III repeat homology; von Willebrand factor type A repeat homology
C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; triple helix
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-110/Domain: fibronectin type III repeat homology <FN3A>
F:156-320/Domain: von Willebrand factor type A repeat homology <VWA1>
F:442-525/Domain: fibronectin type III repeat homology <FN3C>
F:534-614/Domain: fibronectin type III repeat homology <FN3D>
F:623-707/Domain: fibronectin type III repeat homology <FN3E>
F:741-823/Domain: fibronectin type III repeat homology <FN3F>
F:832-914/Domain: fibronectin type III repeat homology <FN3G>
F:922-1009/Domain: fibronectin type III repeat homology <FN3H>
F:1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 8.5%; Score 130; DB 2; Length 1888;
Best Local Similarity 25.0%; Pred. No. 0.05; Indels 26; Gaps 12;
Matches 62; Conservative 44; Mismatches 116;

QY 18 DLYFTLDKSGSV-LHWNHNIYFVEQL--AHKFISPO-LRMSFIVFSTGTMLMKT--E 71
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db 1042 DLVFLVDGWSIGDDNFNKIISFLYSTVGALDKIGPDGTQVAIIQFSDDPRTFEKLNAYK 1101

QY 72 DREIQRGLEELQKVLPGDGYMHGFEFRASEQIYENRQGYRTA--SVIIALTGDGELHE 129
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
Db 1102 TKETLLEAIQOI--AYKGGNTKTKRAIKHARE-VLFTGAGMRKGIPIKVLVITDGRSQD 1158

QY 130 DLFFYSEREANRDL-GAIVYCVGVKDFNETOLARIAD--SKDHVFPVNDGFQALOGII 186
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
Db 1159 DV-----NKVSRMQLDGFSAIGVADADYSELVNIQSKPSERHVFVDD-DFDAFTKIE 1212

QY 187 HSILKKSCEILAAEPSTICAGESFQVVVGRNGFRHARNVDRVLCFSKINDSVTLNEKPF 246
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
Db 1213 DELITFVCETASATCPLVFKDGDKLA-----GFKMMEMFGLVEKEFSALDGVSMPEPGTF 1266

QY 247 SVEDTYLL 254
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
Db 1267 NVYPCYRL 1274

RESULT 11
S42373
hypothetical protein T20G5.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 07-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C:Accession: S42373
R:Smith, A.
submitted to the EMBL Data Library, March 1994
A:Reference number: S42368
A:Accession: S42373
A:Molecule type: DNA
A:Residues: 1-3051 <SMI>
A:Cross-references: EMBL:Z30423; NID:g458479; PID:g458485
C:Genetics:
A:Introns: 96/1; 166/3; 201/2; 355/1; 404/1; 427/2; 794/1; 892/1; 952/2; 1041/1; 1146
C:Superfamily: von Willebrand factor type A repeat homology; EGF homology; fibronectin
```





A:Accession: A44622  
A:Molecule type: mRNA  
A:Residues: 467-546;550-595;752-764 <WOO>  
A:Cross-references: GB:J00185; GB:J00186  
A:Note: the authors translated the codon TAC at 519 as Thr; the nucleic acid translation  
R:Wole, J.E.; Anderson, J.K.; Davison, E.A.; Woods, D.E.  
J. Biol. Chem. 259, 3407-3412, 1984  
A:Title: Complete primary structure for the zymogen of human complement factor B.  
A:Reference number: A20751; MUID:84161997  
A:Accession: A00934  
A:Molecule type: protein; mRNA  
A:Residues: 26-764 <MO>  
A:Cross-references: GB:K01566  
A:Note: nucleic acid translation differs from the sequence shown in having 300-Leu, 328-  
A:Note: 736-Ser was also found  
A:Note: glycosylation sites were determined  
R:Christie, D.L.; Gagnon, J.  
Biochem. J. 209, 61-70, 1983  
A:Title: Amino acid sequence of the Bb fragment from complement factor B. Sequence of th  
A:Reference number: A19188; MUID:83204002  
A:Contents: the final paper in a series documenting the sequence, glycosylation site, an  
A:Accession: A19188  
A:Molecule type: protein  
A:Residues: 260-296, T', 298-764 <CHR>  
R:Campbell, R.D.; Porter, R.R.  
Proc. Natl. Acad. Sci. U.S.A. 80, 4464-4468, 1983  
A:Title: Molecular cloning and characterization of the gene coding for human complement  
A:Reference number: A19947; MUID:83273641  
A:Accession: A19947  
A:Molecule type: DNA  
A:Residues: 346-764 <CAM>  
A:Cross-references: GB:J00125  
A:Accession: B19947  
A:Molecule type: mRNA  
A:Residues: 339-509 <CAL>  
A:Cross-references: GB:J00126; NID:g187723; PIDN:AAA36226.1; PID:g553536  
R:Wu, L.; Morley, B.J.; Campbell, R.D.  
Cell 48, 331-342, 1987  
A:Title: Cell-specific expression of the human complement protein factor B gene: evidenc  
A:Reference number: A25971; MUID:87102880  
A:Accession: B25971  
A:Molecule type: DNA  
A:Residues: 1-99 <WUL>  
A:Cross-references: GB:M15082; NID:g187699; PIDN:AAA59625.1; PID:g553534  
R:Niemann, M.A.; Brown, A.S.; Miller, E.J.  
Biochem. J. 274, 473-480, 1991  
A:Title: The principal site of glycation of human complement Factor B.  
A:Reference number: S14339; MUID:91174758  
A:Accession: S14339  
A:Molecule type: protein  
A:Residues: 270-329 <NTE>  
A:Note: binding site for carbohydrate to lysine under artificial conditions  
R:Morley, B.J.; Campbell, R.D.  
EMBO J. 3, 153-157, 1984  
A:Title: Internal homologies of the Ba fragment from human complement component factor B  
A:Reference number: A44628; MUID:84158524  
A:Accession: A44628  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 16-225, F', 227-259 <MOR>  
R:Schwaebie, W.; Luttig, B.; Sokolowski, T.; Estaller, C.; Weiss, E.H.; Meyer zum Busche  
Immunobiology 188, 221-232, 1993  
A:Title: Human complement factor B: functional properties of a recombinant zymogen of th  
A:Reference number: I54409; MUID:94041399  
A:Accession: I54409  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-764 <RES>  
A:Cross-references: GB:S67310; NID:g452937; PIDN:ADJ3989.1; PID:g4261689  
R:Horiuchi, T.; Kim, S.; Matsumoto, M.; Watanabe, I.; Fujita, S.; Volanakis, J.E.  
Mol. Immunol. 30, 1587-1592, 1993  
A:Title: Human complement factor B: cDNA cloning, nucleotide sequencing, phenotypic conv  
A:Reference number: I57824; MUID:94067177

A:Accession: I57824  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-31, Q', 33-764 <RE2>  
A:Cross-references: GB:I15702; NID:g291921; PIDN:AAA16820.1; PID:g291922  
C:Comment: 292-Cys has a free sulphydryl.  
C:Genetics:  
A:Gene: GDB:BF  
A:Cross-references: GDB:I19726; OMIM:138470  
A:Map position: 6p21.3-6p21.3  
A:Introns: 21/3; 99/3; 346/1; 390/1; 424/1; 470/1; 502/3; 542/1; 593/2; 619/1; 652/3;  
A:Note: the list of introns may be incomplete  
A:Note: gene is located in the major histocompatibility complex, class III region  
C:Complex: complement factor B initially forms an inactive complex with complement fa  
ment factor C3b forming active C3/C5 convertase; Ba is released  
C:Function:  
A:Description: Bb is a serine proteinase; C3/C5 convertase cleaves complement C3 alph  
al  
A:Pathway: complement alternate pathway  
C:Superfamily: complement C2; complement factor H repeat homology; trypsin homology;  
C:Keywords: acute phase; complement alternate pathway; duplication; glycoprotein; hyd  
F:1-25/Domain: signal sequence #status predicted <SIG>  
F:26-764/Product: complement factor B #status experimental <MAT>  
F:26-259/Product: complement factor Ba fragment #status experimental <BAF>  
F:37-98/Domain: complement factor H repeat homology <FH2>  
F:103-158/Domain: complement factor H repeat homology <FH3>  
F:165-218/Domain: complement factor H repeat homology #status experimental <BBF>  
F:260-764/Product: C3/C5 convertase Bb fragment #status experimental <VFA>  
F:268-458/Domain: von Willebrand factor type A repeat homology <TRY>  
F:482-752/Domain: trypsin homology #status atypical <TRY>  
F:37-76,62-98,103-145,131-158,165-205,191-218,478-596,511-527,599-615,656-682,695-725  
F:122,142,285,378/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:259-260/Cleavage site: Arg-Lys (complement factor D) #status experimental  
F:526,576,699/Active site: His, Asp, Ser #status experimental

Query Match	7.4%	Score 113.5	DB 1	Length 764
Best Local Similarity	19.6%	Pred. No. 0.34		
Matches	66	Conservative	63	Mismatches 109; Indels 99; Gaps 17
QY	2	QGGRRDGG-GPA-----	CGGFDLYFILDKSSV-----	LHWNFIYFVQOLAH 45
Db	243	EGVDAEDGCHGPEQQRKRIVLDPGSGNMIVLVDGSDSIGASNFTGAKCLVNLIEKVAS 302		
QY	46	KFTSPQLRMSFIYFSTRGTTLMLKTE----	DREQIRQGLEEL-----	QKVLPGGDTYMHG 97
Db	303	YGVKRP--RYGLVIYATYFKIWKVSEADSSNADWVTKQLNEINVEDHKLKSGTNT-----	355	
QY	98	FERASEQIYYENR-----	OGY-RTASVIALTDG-----	ELHEDLRFYSE 136
Db	356	-KKALQAVYSMMSPDDVPPEGMNRTRHVILMTDGLHNNMGDPITVVIDEIRDLLYIGKD 414		
QY	137	REANRSRDLGAIVCYVG--VKDEFNETOLARTADSKDHVPFVNDGFOALOGIIHSLKSC 194		
Db	415	RKNRPEDYLDVYFVGVLNVQVINALASKKNEQHFVKVKD-MENLEDVYFQIMIDES- 472		
QY	195	TEILLAAPESTICAGESFQVYVVGNGFRHARNDVRLCSFKINDSVTLNEKPFVSVEDYLL 254		
Db	473	-----QSLSLC-----	GMVWEHRKGT-----	YHKOPWQAKISV-- 501
QY	255	CPAPILKEVGKMAALQVSMNDGLSFISSSVITITTHC 291		
Db	502	----IRPSKGHSCMG-----	AVVSEYFVLTAHC 527	
RESULT	15			
A54849				
collagen alpha 1(VII) chain precursor - human				
N:Alternate names: procollagen alpha 1(VII) chain				
C:Species: Homo sapiens (man)				
C:Date: 04-Nov-1994 #sequence_revision 04-Nov-1994 #text_change 20-Sep-1999				
C:Accession: A54849; PH0844; S16316; I56328; I84686				
R:Christiano, A.M.; Greenspan, D.S.; Lee, S.; Uitto, J.				

J. Biol. Chem. 269, 20256-20262, 1994

A:Title: Cloning of human type VII collagen. Complete primary sequence of the alpha1(VII)

A:Reference number: A54849; MUID:94327598

A:Accession: A54849

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-2944 <CHR>

A:Cross-references: NID:987124; PIDN:AAA75438.1; PID:987125

R:Tanaka, T.; Takahashi, K.; Furukawa, F.; Imamura, S.

Biochem. Biophys. Res. Commun. 183, 958-963, 1992

A:Title: Molecular cloning and characterization of type VII collagen cDNA.

A:Reference number: PH0844; MUID:92231392

A:Accession: PH0844

A:Molecule type: mRNA

A:Residues: 'EFR', 340-475, 'RALSTASHSTLCWRATWHPNCRGSHWTRAAACEPCNRPASHRAARAG', 524-528, 'C',

A:Cross-references: DDBJ:D11152; DDBJ:D13694; NID:9453698; PIDN:BA02853.1; PID:9453699

A:Experimental source: keratinocyte

A>Note: the authors translated the codon ACC for residues 394 and 397 as Tyr

R:Parente, M.G.; Chung, L.C.; Ryyanen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Mat

Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991

A:Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.

A:Reference number: S16316; MUID:91334380

A:Accession: S16316

A:Molecule type: mRNA

A:Residues: 815-892, 'E', 894-1439 <PAR>

A:Cross-references: GB:M65158; GB:S49017; NID:9180914; PIDN:AAA96439.1; PID:9180915

A:Experimental source: keratinocyte

R:Gammon, W.R.; Abernethy, M.L.; Padilla, K.M.; Prisyann, P.S.; Cook, M.E.; Wright, J.

J. Invest. Dermatol. 99, 691-696, 1992

A:Title: Noncollagenous (NCL) domain of collagen VII resembles multidomain adhesion prot

A:Reference number: I56328; MUID:93107742

A:Accession: I56328

A:Molecule type: mRNA

A>Status: translated from GB/EMBL/DBJ

A:Residues: 'EFR', 372-517, 'DV', 520-540, 'W', 542-1255 <RES>

A:Cross-references: GB:S51236; NID:9262308; PIDN:AB24637.1; PID:9262309

R:Seltzer, J.L.; Etsen, A.Z.; Bauer, E.A.; Morris, N.P.; Glanville, R.W.; Burgeson, R.E.

J. Biol. Chem. 264, 3822-3826, 1989

A:Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagena

A:Reference number: A30296; MUID:89139437

A:Accession: A30296

A:Molecule type: protein

A:Residues: 'A', 1240-1246, 'G', 1248-1250, 'XE', 1253-1255, 'Q', 1257, 'E', 2032, 'C', 2034-2041;

A>Note: two reported peptides cannot be reliably located

R:Greenspan, D.S.

Hum. Mol. Genet. 2, 273-278, 1993

A:Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous

A:Reference number: I48103; MUID:93271985

A:Accession: I48103

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 2395-2871, 'S', 2873-2944 <RES>

A:Cross-references: GB:I06862; NID:9388713; PIDN:AAA89196.1; PID:9388714

R:Christiano, A.M.; Ryyanen, M.; Gittto, J.

Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994

A:Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --> Ser subs

A:Reference number: A55255; MUID:94224777

A:Contents: annotation

C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C

ed and subsequently O-glycosylated.

C:Genetics:

A:Gene: GDB:COL7A1; GDB:EBD1; EB

A:Cross-references: GDB:128750; OMIM:120120

A:Map position: 3p21.3-3p21.3

A>Note: defects in this gene can result in dominant and recessive dystrophic epidermolys

C:Complex: type VII collagen is probably a homotrimer

C:Function:

A:Description: structural component of extracellular polymer associated with anchoring f

C:Superfamily: unassigned collagens; animal Kunitz-type proteinase inhibitor homolog; f

C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxyllysine; hydroxyproli

F:1-16/Domain: signal sequence #status predicted <SIG

F:17-2944/Product: collagen alpha 1(VII) chain #status predicted <NAT>

F:17-1253/Domain: amino-terminal nonhelical #status predicted <NC1>  
F:36-201/Domain: von Willebrand factor type A repeat homology <VWA1>  
F:231-318/Domain: fibronectin type III repeat homology <FN1>  
F:327-413/Domain: fibronectin type III repeat homology <FN2>  
F:414-502/Domain: fibronectin type III repeat homology <FN3>  
F:508-593/Domain: fibronectin type III repeat homology <FN4>  
F:598-683/Domain: fibronectin type III repeat homology <FN5>  
F:686-771/Domain: fibronectin type III repeat homology <FN6>  
F:776-862/Domain: fibronectin type III repeat homology <FN7>  
F:864-952/Domain: fibronectin type III repeat homology <FN8>  
F:954-1045/Domain: fibronectin type III repeat homology <FN9>  
F:1052-1219/Domain: von Willebrand factor type A repeat homology <VWA2>  
F:1170-1172/Region: cell attachment (R-G-D) motif  
F:1189-1253/Region: cysteine/proline-rich  
F:1254-2783/Region: interrupted helical  
F:1334-1336/Region: cell attachment (R-G-D) motif  
F:2008-2010/Region: cell attachment (R-G-D) motif  
F:2553-2555/Region: cell attachment (R-G-D) motif  
F:2784-2944/Domain: carboxyl-terminal nonhelical #status predicted <NC2>  
F:2876-2929/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>  
F:337-786-1109/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:2167-2176,2185-2188,2664,2667,2673/Modified site: 4-hydroxyproline (Pro) #status ex  
F:2625,2631/Modified site: 5-hydroxylysine (Lys) #status experimental  
F:2625,2631/Binding site: carbohydrate (Lys) (covalent) #status experimental  
F:2634-2802-2804/Disulfide bonds: interchain #status predicted

[illegible]

Search completed: August 9, 2002, 10:34:15  
Job time: 254 sec

us-09-970-076-2\_copy\_27\_321.rpr

Fri Aug 9 10:57:02 2002

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:32:10 ; Search time 119.13 seconds  
(without alignments)  
275.051 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_27\_321

Perfect score: 1536

Sequence: 1 GCGGRRDGGPACYGCFDLY.....GLSFSSSVITTHCSDGS 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802.\*  
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1536	100.0	403	22	AAE01439 Human gene 4 encod
2	1524	99.2	403	22	AAE01469 Human gene 4 encod
3	1520	99.0	333	21	AAE01422 Human TANGO 197.
4	1263	82.2	297	22	AAE01422 Human polypeptide
5	966	62.9	218	22	AAE01422 Human protein sequ
6	791.5	51.5	488	21	AAE01456 A human TANGO 216
7	790.5	51.5	488	21	AAE01447 Amino acid sequenc
8	790.5	51.5	488	21	AAE01455 A human TANGO 216
9	787.5	51.3	488	21	AAE01457 A human TANGO 216
10	777.5	50.6	587	22	AAU19662 Human novel extrac
11	770	50.1	487	21	AAE01448 Amino acid sequenc

12	770	50.1	487	21	AAE01458 A murine TANGO 216
13	769	50.1	487	21	AAE01460 A murine TANGO 216
14	767	49.9	487	21	AAE01459 A murine TANGO 216
15	685	44.6	381	21	AAE01428 Murine TANGO 197.
16	205	13.3	93	22	AAE01430 Human immune/haema
17	166	10.8	1155	16	AAE01431 Mouse alpha-d subu
18	166	10.8	1155	18	AAE01432 Mouse beta 2 integ
19	166	10.8	1155	19	AAE01433 Mouse alpha-d #1.
20	166	10.8	1155	19	AAE01434 Mouse beta-integri
21	166	10.8	1155	19	AAE01435 Mouse alpha d poly
22	166	10.8	1155	20	AAE01436 Mouse alphas prote
23	166	10.8	1155	21	AAE01437 Mouse alpha-d prot
24	166	10.8	1161	16	AAE01438 AA078168
25	166	10.8	1161	18	AAE01439 AA0723061
26	166	10.8	1161	19	AAE01440 AA072836
27	166	10.8	1161	19	AAE01441 AA065103
28	166	10.8	1161	19	AAE01442 AA060003
29	166	10.8	1161	20	AAE01443 AA073347
30	166	10.8	1161	21	AAE01444 AA073373
31	159.5	10.4	1151	18	AAE01445 AA023059
32	159.5	10.4	1151	19	AAE01446 AA072834
33	159.5	10.4	1151	19	AAE01447 AA065101
34	159.5	10.4	1151	19	AAE01448 AA060001
35	159.5	10.4	1151	20	AAE01449 AA073344
36	159.5	10.4	1151	21	AAE01450 AA073371
37	159.5	10.4	1161	16	AAE01451 AA078169
38	159.5	10.4	1161	18	AAE01452 AA023062
39	159.5	10.4	1161	19	AAE01453 AA072824
40	159.5	10.4	1161	19	AAE01454 AA065104
41	159.5	10.4	1161	19	AAE01455 AA060004
42	159.5	10.4	1161	20	AAE01456 AA073345
43	159.5	10.4	1161	21	AAE01457 AA073374
44	158.5	10.3	1151	16	AAE01458 AA078179
45	156	10.2	1161	16	AAE01459 AA078166

#### ALIGNMENTS

RESULT 1

AAE01439

ID AAE01439 standard; Protein: 403 AA.

XX AC AAE01439;

XX AC AAE01439;

DT 17-JUL-2001 (first entry)

XX

DE Human gene 4 encoded secreted protein HWLFR02, SEQ ID NO:94.

Human; secreted protein; proliferative disorder; cancer; tumour;  
foetal abnormality; developmental abnormality; haematopoietic disorder;  
immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
inflammation; allergy; neurological disorder; Alzheimer's disease;  
Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
cardiovascular disorder; angioneurotic disorder; kidney disorder;  
gastrointestinal disorder; pregnancy-related disorder;  
endocrine disorder; infection; wound healing; vulnery;  
cell culture; chemotaxis; food additive; gene therapy;  
binding partner identification; chromosome 19.

OS Homo sapiens.

XX

XX Key Location/Qualifiers

FT Peptide /label= Signal\_peptide

FT Protein 1..27

FT Protein 28..403

XX /note= "Mature human secreted protein"

XX WO200134626-A1.

XX

XX 17-MAY-2001.

XX

XX

F		01-NOV-2000; 2000WO-US30045.	
R	X	05-NOV-1999; 99US-0163581.	
R	X	30-JUN-2000; 2000US-0215133.	
X	A	(HUMA-) HUMAN GENOME SCI INC.	
I	X	Ruben SM, Komatsoulis GA, Moore PA, Birse CE, Ni J;	
X	X	WPI; 2001-308778/32.	
S	X	N-PSDB; AAD05303.	
X	T	New nucleic acid molecules encoding 28 human secreted proteins for	
T	T	diagnosing, preventing, treating or ameliorating medical conditions and	
T	T	used as food additives or preservatives -	
X	S	Claim 11; Page 485-486; 562pp; English.	
X	X	AAD05300-AAD05379 represent cDNAs corresponding to 28 human secreted	
X	X	protein genes, and AAE01436-AAE01513 represent the proteins they encode.	
C	C	AAE01514-AAE01544 represent human secreted protein fragments or variants.	
C	C	The genes and their secreted proteins are useful for preventing,	
C	C	treatment or ameliorating medical conditions, e.g., by protein or gene	
C	C	therapy. Pathological conditions can be diagnosed by determining the	
C	C	amount of the new protein in a sample or by determining the presence of	
C	C	mutations in the new genes. Specific uses are described for each of the	
C	C	28 genes, based on the tissues in which they are most highly expressed,	
C	C	and include developing products for the diagnosis or treatment of	
C	C	proliferative disorders, cancer, tumours, foetal and developmental	
C	C	abnormalities, haematopoietic disorders, diseases of the immune system,	
C	C	AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,	
C	C	allergies, neurological disorders (e.g., Alzheimer's disease,	
C	C	Parkinson's disease), cognitive disorders, schizophrenia, asthma,	
C	C	skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,	
C	C	cardiovascular disorders, angiotensin-related disorders, endocrine	
C	C	gastrointestinal disorders, pregnancy-related disorders, endocrine	
C	C	disorders, and infectious. The proteins can also be used to aid wound	
C	C	healing and epithelial cell proliferation, to prevent skin aging due to	
C	C	sunburn, to maintain organs before transplantation, for supporting cell	
C	C	culture of primary tissues, to regenerate tissues, to identify their	
C	C	cognate ligands or binding partners, and in chemotaxis, and can be used	
C	C	as a food additive or preservative to modify storage properties.	
C	C	Antibodies specific for a protein of the invention can be used in	
C	C	alleviating symptoms associated with the disorders mentioned above, and	
C	C	in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked	
C	C	immunosorbent assay (ELISA). The present sequence represents a human	
C	C	secreted protein of the invention.	
X	SQ	Sequence 403 AA;	
		Query Match 100.0%; Score 1536; DB 22; Length 403;	
		Best Local Similarity 100.0%; Pred. No. 5.3e-157;	
		Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY		1 GGGRREDGGPACYGFDLYFLDKSGSVLHHNNEYFYVEQLAHKFISPOLRMSPFVS 60	
Db			
		27 gggrrredggpacygfdlyflldksvglhhwnelyyveqlahkfispqrlmsfivs 86	
QY		61 TRGTTLMKLTEDREQIRGLEEKVLPGDGYMHGCFERASQIYYENRQGYRTASVII 120	
Db			
		87 trgttlmkltedreqirgleelkvlpgdtymhgferaseqiyyenrqqyrtasvii 146	
QY		121 ALTDGELHDLFFYSEREANRSRLGAIVYCVGVKDFNETQLARIADSKDHVPVNDGFQ 180	
Db			
		147 altdgelhdelffyserearnsrdlgaivycvgvkdfnetqlariadskdhvipndgfq 206	
QY		181 ALQGIHSILKSCIEILAEPPTICAGESFQVVVRNGFRHARNVDRLVLCFSKINDSVT 240	
Db			
		207 algqlihsilkkscieilaepsticagesfqvvvrngfrharnvdrlvcfskindsvt 266	
QY		241 LNEKPFVEDTYLLCPAPILKEVGMAALQVNDGLSFSSVLIITTHCSOGS 295	

Db 267 lnekpfsvedtyllcpapilkevgmkaalqvsmndglslsfssvliitthcsogds 321

RESULT 2  
ID AAE01469 standard; Protein; 403 AA.  
XX AC AAE01469;  
XX DT 17-JUL-2001 (first entry)  
XX DE Human gene 4 encoded secreted protein HWLER02, SEQ ID NO:125.  
DE KW Human; secreted protein; proliferative disorder; cancer; tumour;  
KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
KW immune system allergy; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergic; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
KW gastrointestinal disorder; pregnancy-related disorder;  
KW endocrine disorder; infection; wound healing; vulnerability;  
KW cell culture; chemotaxis; food additive; gene therapy;  
KW binding partner identification; chromosome 19.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
FT Peptide 1..27 /label= Signal\_peptide  
FT Protein 28..403 /note= "Mature human secreted protein"  
FT Misc-difference 175 /label= Unknown  
FT Misc-difference 320 /note= "Encoded by GKT"  
FT Misc-difference 331 /label= Unknown  
FT Misc-difference 368 /note= "Encoded by KTC"  
FT Misc-difference 368 /label= Unknown  
FT Misc-difference 368 /note= "Encoded by WGC"  
XX WO200134626-A1.  
XX PD 17-MAY-2001.  
XX PF 01-NOV-2000; 2000WO-US30045.  
XX PR 05-NOV-1999; 99US-0163581.  
XX PR 30-JUN-2000; 2000US-0215133.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX XX Ruben SM, Komatsoulis GA, Moore PA, Birse CE, Ni J;  
XX WPI; 2001-308778/32.  
XX N-PSDB; AAD05334.  
XX New nucleic acid molecules encoding 28 human secreted proteins for  
XX diagnosing, preventing, treating or ameliorating medical conditions and  
XX used as food additives or preservatives -  
XX Claim 11; Page 505-506; 562pp; English.  
XX AAD05300-AAD05379 represent cDNAs corresponding to 28 human secreted  
XX protein genes, and AAE01436-AAE01513 represent the proteins they encode.  
XX AAE01514-AAE01544 represent human secreted protein fragments or variants.  
XX The genes and their secreted proteins are useful for preventing,  
XX treatment or ameliorating medical conditions, e.g., by protein or gene  
XX therapy. Pathological conditions can be diagnosed by determining the  
XX amount of the new protein in a sample or by determining the presence of  
XX mutations in the new genes. Specific uses are described for each of the  
XX 28 genes, based on the tissues in which they are most highly expressed,  
XX and include developing products for the diagnosis or treatment of  
XX proliferative disorders, cancer, tumours, foetal and developmental  
XX abnormalities, haematopoietic disorders, diseases of the immune system,  
XX AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
XX allergies, neurological disorders (e.g., Alzheimer's disease,  
XX Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
XX skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
XX cardiovascular disorders, angiotensin-related disorders, endocrine  
XX gastrointestinal disorders, pregnancy-related disorders, endocrine  
XX disorders, and infectious. The proteins can also be used to aid wound  
XX healing and epithelial cell proliferation, to prevent skin aging due to  
XX sunburn, to maintain organs before transplantation, for supporting cell  
XX culture of primary tissues, to regenerate tissues, to identify their  
XX cognate ligands or binding partners, and in chemotaxis, and can be used  
XX as a food additive or preservative to modify storage properties.  
XX Antibodies specific for a protein of the invention can be used in  
XX alleviating symptoms associated with the disorders mentioned above, and  
XX in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
XX immunosorbent assay (ELISA). The present sequence represents a human  
XX secreted protein of the invention.

CC amount of the new protein in a sample or by determining the presence of  
CC mutations in the new genes. Specific uses are described for each of the  
CC 28 genes, based on the tissues in which they are most highly expressed,  
CC and include developing products for the diagnosis or treatment of  
CC proliferative disorders, cancer, tumours, foetal and developmental  
CC abnormalities, haematopoietic disorders, diseases of the immune system,  
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
CC allergies, neurological disorders (e.g., Alzheimer's disease,  
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
CC cardiovascular disorders, angio-genic disorders, kidney disorders,  
CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
CC disorders, and infections. The proteins can also be used to aid wound  
CC healing and epithelial cell proliferation, to prevent skin aging due to  
CC sunburn, to maintain organs before transplantation, for supporting cell  
CC culture of primary tissues, to regenerate tissues, to identify their  
CC cognate ligands or binding partners, and in chemotaxis, and can be used  
CC as a food additive or preservative to modify storage properties.  
CC Antibodies specific for a protein of the invention can be used in  
CC alleviating symptoms associated with the disorders mentioned above, and  
CC in diagnostic immunoassays (e.g., radioimmunoassay or enzyme linked  
CC immunosorbent assay (ELISA)). The present sequence represents a human  
CC secreted protein of the invention.

XX Sequence 403 AA;

Query Match 99.28; Score 1524; DB 22; Length 403;  
Best Local Similarity 99.3%; Pred. No. 1.le-155;  
Matches 293; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGGRRDGGPACYGFDLYFLDKSGSVLHWNHNIYFVEQLAHKFISPOLRMSFIVFS 60  
Db 27 99grrdggpacyggfdlyfldksgsvlhwnhniyfyveqlahkfispqlrmsfivfs 86  
Qy 61 TRGTTMLKLTEDREQIROGLEELQKVLPGDGYMHGFERASEQIYYENRGYRTASVII 120  
Db 87 trgttlmkltdreqirgleelqkvlpggdtymhgefaseraseqiyyenrgyrtasvii 146  
Qy 121 ALTDGELHEDLFFYSEREANRSDLGAIYCVGVKDFNETOLARIADSKDHVFPVNDGFQ 180  
Db 147 altdgelhedlffysereansrdlgaiycvgvkdfnetqlariadskdhvfpvndgfq 206  
Qy 181 ALQGIHSILKKSCIEILAAEPSTICAGESFQVVVGRNGFRHARNVDRVLCFSKINDSVT 240  
Db 207 alqglihsilkkscielaaepsticagesfqvvvrgngfgharnvdrvlcsfkindsvt 266  
Qy 241 LNEKPFSEVEDYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVIITTHCS DGS 295  
Db 267 lnekpfsvedyllcpapilkevgmkaalqvsmndglsfissviiittthcsdxs 321

RESULT 3  
AAB01422  
ID AAB01422 standard; Protein: 333 AA.

XX AAB01422;

XX 20-OCT-2000 (first entry)

XX Human TANGO 197.

XX TANGO: 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;  
KW graft versus-host diseases; rheumatoid arthritis; psoriasis;  
KW inflammatory bowel disease; septic shock; ulcerative colitis;  
KW Crohn's disease; chronic myelogenous leukemia; cancer; liver  
KW disease; Hodgkin's disease; osteoarthritis; Lyme's disease;  
KW cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;  
KW systemic lupus erythematosus; transgenic animal; diagnosis;  
KW prognosis; prophylactic; therapeutic; human.

XX Homo saplens.

OS

PN WO2000039284-A1.  
XX  
PD 06-JUL-2000.  
XX  
PF 23-DEC-1999; 99WO-US31025.  
XX  
PR 30-DEC-1998; 98US-0223546.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Holtzman DA;  
XX  
DR WPI: 2000-465743/40.  
DR N-PSDB; AAA47455.  
XX  
PT Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,  
PT 224 and 239 polypeptides useful for the treatment of asthma, rheumatoid  
PT arthritis, psoriasis and autoimmune diseases  
XX  
PS Claim 8; Fig 4; 209pp; English.  
XX  
CC Nucleic acids encoding TANGO polypeptides are useful as modulating  
CC agents for regulating cellular processes like asthma, graft  
CC versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory  
CC bowel disease, septic shock, ulcerative colitis, Crohn's disease,  
CC chronic myelogenous leukemia, cancer, liver disease, Hodgkin's  
CC disease, osteoarthritis, Lyme's disease, cachexia and autoimmune  
CC diseases e.g. myasthenia gravis, autoimmune diabetes and systemic  
CC lupus erythematosus. The nucleic acids are also useful for producing  
CC transgenic animals and the TANGO polypeptides themselves. Partial  
CC TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in  
CC forensic biology, for diagnostic assays, prognostic assays,  
CC pharmacogenomics and for monitoring clinical trials. TANGO  
CC polypeptides are suitable for both prophylactic and therapeutic  
CC methods for treating a subject at risk of a disorder or having a  
CC disorder associated with aberrant TANGO expression. A wide range  
CC of cellular disorders can be treated.  
XX  
SQ Sequence 333 AA;

Query Match 99.0%; Score 1520; DB 21; Length 333;  
Best Local Similarity 100.0%; Pred. No. 2.le-155;  
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGRRDGGPACYGFDLYFLDKSGSVLHWNHNIYFVEQLAHKFISPOLRMSFIVFS 60  
Db 27 99grrdggpacyggfdlyfldksgsvlhwnhniyfyveqlahkfispqlrmsfivfs 86  
Qy 61 TRGTTMLKLTEDREQIROGLEELQKVLPGDGYMHGFERASEQIYYENRGYRTASVII 120  
Db 87 trgttlmkltdreqirgleelqkvlpggdtymhgefaseraseqiyyenrgyrtasvii 146  
Qy 121 ALTDGELHEDLFFYSEREANRSDLGAIYCVGVKDFNETOLARIADSKDHVFPVNDGFQ 180  
Db 147 altdgelhedlffysereansrdlgaiycvgvkdfnetqlariadskdhvfpvndgfq 206  
Qy 181 ALQGIHSILKKSCIEILAAEPSTICAGESFQVVVGRNGFRHARNVDRVLCFSKINDSVT 240  
Db 207 alqglihsilkkscielaaepsticagesfqvvvrgngfgharnvdrvlcsfkindsvt 266  
Qy 241 LNEKPFSEVEDYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVIITTHCS 292  
Db 267 lnekpfsvedyllcpapilkevgmkaalqvsmndglsfissviiittthcs 318

RESULT 4  
AAM38976  
ID AAM38976 standard; Protein: 297 AA.

XX AAM38976;

XX 22-OCT-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 2121.

XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

XX KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

KW leukaemia.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX XX 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-052317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

DR N-PSDB; AAI58132.

XX DR Novel nucleic acids and polypeptides, useful for treating disorders

XX PT such as central nervous system injuries -

XX PS Example 4; SEQ ID NO 2121; 10078pp; English.

XX XX The invention relates to human nucleic acids (AA157798-AA161369) and

XX CC the encoded polypeptides (AAM38642-AAAM42213) with nootropic,

XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful

XX CC in gene therapy. A composition containing a polypeptide or polynucleotide

XX CC of the invention may be used to treat diseases of the peripheral nervous

XX CC system, such as peripheral nervous injuries, peripheral neuropathy and

XX CC localised neuropathies and central nervous system diseases, such as

XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

XX CC utilisation of the activities such as: Immune system suppression,

XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

XX CC assays for receptor activity, arthritis and inflammation, leukaemias and

XX CC C.N.S disorders.

XX CC Note: The sequence data for this patent did not form part of the printed

XX CC specification.

XX SQ Sequence 297 AA;

Query Match 82.2%; Score 1263; DB 22; Length 297;

Best Local Similarity 99.2%; Pred. No. 1e-127; 0; Indels 0; Gaps 0;

Matches 241; Conservative 2; Mismatches 0;

QY 1 GGGRRDGGPACVGGFDLYFLDKSGSVLHHWNEIYFVEQLAHKFIQPLRMSFIVFS 60

Db 27 gggrrdggpacygggfdlyflldksgsvlhhwnelyfveqlahkfispqlrmfivfs 86

QY 61 TRGTTMLKLTEDRQIFRQGLEELQKVLPGDGYMHGCFERASQIYYENRQGYRTASVII 120

Db 87 trgttlmkltdreqirggleelqkvlpvgdgytmhgefaseraseqiyenrggyrtasvii 146

QY 121 ALTDGELHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQLARIADSKDHVPVNDGFG 180

Db 147 altgelhedlffysereanrsgldgalvycvgvkdnetqlariadskdhvfpvndgfg 206

QY 181 ALQGIHHSILKSCIEIILAAEPSVICAGESFQVVVRGNGRPHARNVDRLVLCSEKINDSVT 240

Db 207 alqgihsilkkscieilaepsticagesfqvvvrgngfhrnrvlcsfkindsvt 266

QY 241 LNE 243

Db 267 lsk 269

RESULT 5

AAB92985

ID AAB92985 standard; Protein; 218 AA.

XX AC AAB92985;

XX XX 26-JUN-2001 (first entry)

XX DT Human protein sequence SEQ ID NO:11706.

XX DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX KW Homo sapiens.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX XX 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX XX (HELI-) HELIX RES INST.

XX XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX PT Primer sets for synthesizing polynucleotides, particularly the 5602

XX PT full-length cDNAs defined in the specification, and for the detection

XX PT and/or diagnosis of the abnormality of the proteins encoded by the

XX PT full-length cDNAs -

XX XX Claim 8; SEQ ID 11706; 2537pp + CD ROM; English.

XX PS The present invention describes primer sets for synthesising 5602

XX CC full-length cDNAs defined in the specification. Where a primer set

XX CC comprises: (a) an oligo-drimer and an oligonucleotide complementary

XX CC to the complementary strand of a polynucleotide which comprises one of

XX CC the 5602 nucleotide sequences defined in the specification, where the

XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX CC of an oligonucleotide comprising a sequence complementary to the

XX CC complementary strand of a polynucleotide which comprises a 5'-end

XX CC sequence and an oligonucleotide comprising a sequence complementary to a

XX CC polynucleotide which comprises a 3'-end sequence, where the

XX CC oligonucleotide comprises at least 15 nucleotides and the combination in

XX CC the 5'-end sequence/3'-end sequence is selected from those defined in

XX CC the specification. The primer sets can be used in antisense therapy and

XX CC in gene therapy. The primers are useful for synthesising polynucleotides,

XX CC particularly full-length cDNAs. The primers are also useful for the

XX CC detection and/or diagnosis of the abnormality of the proteins encoded by

XX CC the full-length cDNAs. The primers allow obtaining of the full-length

XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

XX CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632



CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

SQ Sequence 218 AA;  
  
Query Match 62.9%; Score 966; DB 22; Length 218;  
Best Local Similarity 98.9%; Pred. No. 7.9e-96;  
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 54 MSFVSTRGTTLMKLTEDRQIRQGLEELQKVLPGDGYHHEGFERASEQIYYENRGY 113  
DB 1 msfivstrgttlmkltedreqirggleelqkvlpgdgyhhegferasedqiiyenrggy 60  
  
QY 114 RTASVITIALTGELHEDLFFYSERANRSRLGALVYCVGVKDFNETOLARIADSKDHVF 173  
DB 61 rtasvialtdgelhedlffysereansrldgalvycvgvkdfnetqlariadskdhvf 120  
  
QY 174 PVNDGFQALQGIHLSILKSCIEILAAEPSTICAGESQVVRGNGFRHARNVDRLVCSF 233  
DB 121 pvndgfqalqgihsilkkscieilaepsticagesqvvvrgngfrharnvdrvlcsf 180  
  
QY 234 KINDSVTLINE 243  
DB 181 kindsvtlisk 190

RESULT 6  
AAB18456  
ID AAB18456 standard; Protein; 488 AA.  
XX AAB18456;  
XX 15-JAN-2001 (first entry)  
XX A human TANGO 216 polypeptide clone.  
XX TANGO 266; TANGO 216;; TANGO 262; TANGO 267;  
KW cellular proliferation; cellular differentiation; cellular adhesion;  
KW von Willebrand factor-associated disorder; cell trafficking; cancer;  
KW hematopoietic associated disease; atelectasis; pulmonary congestion;  
KW edema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;  
KW intestinal disorder; spleen associated disease; renal disorder;  
KW cardiovascular disorder; ischemic heart disease; hydrocephalus;  
KW brain herniation; atrogenic disease; inflammation; meningitis;  
KW Alzheimer's disease; cerebral toxoplasmosis; Parkinson's disease;  
KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.  
XX Homo sapiens.  
XX WO200052022-A1.  
XX 08-SEP-2000.  
XX 01-MAR-2000; 2000WO-US05226.  
XX 01-MAR-1999; 99US-0122458.  
XX (MILL-) MILLENNIUM PHARM INC.  
XX Barnes TM, Holtzman DA, Sharp JD, Fraser CC;  
XX WPI; 2000-579269/54.  
XX N-PSDB; AAA75158.  
XX Novel human and murine secreted proteins designated TANGO 216, 261,  
XX 262, 266 and 267 useful as modulating agents of cellular processes,  
XX e.g. for treating cancer -  
XX Disclosure; Page -; 175pp; English.  
XX AAB18455-57 represent human TANGO 216 proteins. The specification also  
XX describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO

CC polypeptides can be used to modulate cellular proliferation, modulate  
CC cellular differentiation and/or modulate cellular adhesion. The  
CC proteins can be used to treat any von Willebrand factor-associated  
CC disorder, regulate extracellular matrix structuring, cellular adhesion,  
CC and cell trafficking and/or migration, modulate cellular interactions,  
CC modulate cell adhesion in proliferative disorders, such as cancer,  
CC that appear in the bone marrow, and leukocytes, treat bone marrow, blood  
CC and hematopoietic associated diseases and disorders, atelectasis,  
CC pulmonary congestion or oedema, emphysema, chronic bronchitis, bronchial  
CC asthma and bronchiectasis, intestinal disorders, spleen associated  
CC diseases, modulate renal disorders, treat cardiovascular disorders such  
CC as ischemic heart disease, modulate the proliferation, differentiation,  
CC and/or function of bone and cartilage cells and to treat bone and/or  
CC cartilage associated diseases or disorder. They may also be used to  
CC treat disorders associated with the ovaries, and cerebral oedema,  
CC hydrocephalus, brain herniations, atrogenic disease, inflammations,  
CC bacterial and viral meningitis, Alzheimer's disease, cerebral  
CC toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers,  
CC hydrocephalus and encephalitis, and treat hepatic disorders.  
CC note: the present sequence does not appear in the specification; it was  
CC created using information provided.

XX SQ Sequence 488 AA;  
  
Query Match 51.5%; Score 791.5; DB 21; Length 488;  
Best Local Similarity 53.2%; Pred. No. 2e-76; 81; Indels 3; Gaps 2;  
Matches 157; Conservative 54; Mismatches 54;  
  
QY 1 GQGG-RREDGGPACYGDFLYFLDKSGSVLHHNIEIYFVEQLAKHFIQPOLRMSFIVF 59  
DB 26 gpggllraqdpsccrafdlyfldksgsvannwlelynfvqlaerfvspeimrfivf 85  
  
QY 60 STRGTTLMKLTEDRQIRQGLEELQKVLPGDGYHHEGFERASEQIYYENRGYRTASVI 119  
DB 86 ssgatillptgdrqiskgledlkrvspvgetyhegliklaneql--qkagglktssif 143  
  
QY 120 TALTDGELHEDLFFYSERANRSRLGALVYCVGVKDFNETOLARIADSKDHVFPVNDGF 179  
DB 144 ialtdgkldgvlpsyaekaisrslgasvycvgldefeqadleradskeqvfvpkqgfv 203  
  
QY 180 QALQGIHLSILKSCIEILAAEPSTICAGESQVVRGNGFRHARNVDRLVCSFKINDSV 239  
DB 204 qalkgilnsilaqscetleilqpsvcvgeefqivlgslgrngslgrngsvlctytvnevy 263  
  
QY 240 TLNEPFSVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFISSSVITITTHCSDG 294  
DB 264 ttsvxpvsqvqnsmlcpapilnkagetedvsvfnggksvsgslivtatecsng 318

RESULT 7  
AAB18447  
ID AAB18447 standard; Protein; 488 AA.  
XX AAB18447;  
XX 15-JAN-2001 (first entry)  
XX Amino acid sequence of human TANGO 216 polypeptide.  
XX TANGO 266; TANGO 216;; TANGO 261; TANGO 262; TANGO 267;  
KW cellular proliferation; cellular differentiation; cellular adhesion;  
KW von Willebrand factor-associated disorder; cell trafficking; cancer;  
KW hematopoietic associated disease; atelectasis; pulmonary congestion;  
KW edema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;  
KW intestinal disorder; spleen associated disease; renal disorder;  
KW cardiovascular disorder; ischemic heart disease; hydrocephalus;  
KW brain herniation; atrogenic disease; inflammation; meningitis;  
KW Alzheimer's disease; cerebral toxoplasmosis; Parkinson's disease;  
KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.  
XX Homo sapiens.



CC and hematopoietic associated diseases and disorders, atelectasis,  
CC pulmonary congestion or oedema, emphysema, chronic bronchitis, bronchial  
CC asthma and bronchiectasis, intestinal disorders, spleen associated  
CC diseases, modulate renal disorders, treat cardiovascular disorders such  
CC as ischemic heart disease, modulate the proliferation, differentiation,  
CC and/or function of bone and cartilage cells and to treat bone and/or  
CC cartilage associated diseases or disorder. They may also be used to  
CC treat disorders associated with the ovaries, and cerebral oedema,  
CC hydrocephalus, brain herniations, iatrogenic disease, inflammations,  
CC bacterial and viral meningitis, Alzheimer's disease, cerebral  
CC toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers,  
CC hydrocephalus and encephalitis, and treat hepatic disorders.  
CC note: the present sequence does not appear in the specification; it was  
CC created using information provided.  
XX  
SQ Sequence 488 AA;

Query Match 51.5%; Score 790.5; DB 21; Length 488;  
Best Local Similarity 53.2%; Pred. No. 2.5e-76;  
Matches 157; Conservative 54; Mismatches 81; Indels 3; Gaps 2;

QY 1 GGGG-RREDGGPACGCGFDLYFILDKSGSVLHHWNEIYFVQLAHKFISPOLRMSFIVF 59  
DB 26 gpqgilraqeqpsccrafdlyfvlksgsvannwidiynfvqqlaerfvspeimrlsflvf 85  
QY 60 STRGTTLMKLTREDQIROGLEELQKVLPGDGYMHGFERASEQIYYENROGYRTASVI 119  
DB 86 ssqatiilptdgrkiskgledlkrvspvgetyhegiklaneql--qkagglktsll 143  
QY 120 IALTDELHEDLFFYSERANRSDLGAIYVCVGVKDFNETQARIADSKDHVPVNDGF 179  
DB 144 ialtdgklglvpsyaekaisrsgasvycgvldfegaqleriadskeqfvpkggf 203  
QY 180 QALQGIHHSILKSCIEILAAEPSITCAGESQVQVVRGNGFHRHNRVDRVLSFKINDSV 239  
DB 204 qalkgilsilagscteilqlpssvcvgeefqivlsgrfmlgsrngsvlctytvety 263  
QY 240 TUNEPFVSVEDTYLLCPAPILKEVGKMAALOVSMDGLSFISSSVITTTTHCSGDG 294  
DB 264 ttsvkvpsvqlnsmicpapilnkagetldsvsfnggksvsgslivtatecsng 318

Query Match 51.3%; Score 787.5; DB 21; Length 488;  
Best Local Similarity 52.9%; Pred. No. 5.3e-76;  
Matches 156; Conservative 55; Mismatches 81; Indels 3; Gaps 2;

QY 1 GGGG-RREDGGPACGCGFDLYFILDKSGSVLHHWNEIYFVQLAHKFISPOLRMSFIVF 59  
DB 26 gpqgilraqeqpsccrafdlyfvlksgsvannwidiynfvqqlaerfvspeimrlsflvf 85  
QY 60 STRGTTLMKLTREDQIROGLEELQKVLPGDGYMHGFERASEQIYYENROGYRTASVI 119  
DB 86 ssqatiilptdgrkiskgledlkrvspvgetyhegiklaneql--qkagglktsll 143  
QY 120 IALTDELHEDLFFYSERANRSDLGAIYVCVGVKDFNETQARIADSKDHVPVNDGF 179  
DB 144 ialtdgklglvpsyaekaisrsgasvycgvldfegaqleriadskeqfvpkggf 203  
QY 180 QALQGIHHSILKSCIEILAAEPSITCAGESQVQVVRGNGFHRHNRVDRVLSFKINDSV 239  
DB 204 qalkgilsilagscteilqlpssvcvgeefqivlsgrfmlgsrngsvlctytvety 263  
QY 240 TUNEPFVSVEDTYLLCPAPILKEVGKMAALOVSMDGLSFISSSVITTTTHCSGDG 294  
DB 264 ttsvkvpsvqlnsmicpapilnkagetldsvsfnggksvsgslivtatecsng 318

RESULT 10  
AAU19662  
ID AAU19662 standard; Protein; 587 AA.  
XX  
AC AAU19662;  
XX

Fri Aug 9 10:57:01 2002

DT	06-DEC-2001	(first entry)	
XX	Human novel extracellular matrix protein, Seq ID No 312.		
DE	Human: secreted extracellular matrix protein; immunomodulatory;		
XX	Anti-HIV; antianemic; antirheumatic; antisclerotic; cardiac; vascular;		
KW	cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytostatic;		
KW	antialzheimers; immune/autoclone disease; HIV infection; anaemia;		
KW	human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;		
KW	cancers; hyperproliferative disorder; breast neoplasm; melanoma;		
KW	Sézary syndrome; Gaucher's disease; neurological diseases;		
KW	Alzheimer's disease; Parkinson's disease; cardiovascular disorder;		
KW	cardiac arrest; tachycardia; angina; infection; corneal infections;		
KW	wound healing; immunogen; gene therapy; antisense; food additive.		
XX	Homo sapiens.		
OS	WO20015368-A1.		
XX	02-AUG-2001.		
PN	17-JAN-2001; 2001WO-US01348.		
PD	31-JAN-2000; 2000US-0179065.		
PF	04-FEB-2000; 2000US-0180628.		
XX	24-FEB-2000; 2000US-0184664.		
XX	02-MAR-2000; 2000US-0186350.		
XX	16-MAR-2000; 2000US-0189874.		
XX	17-MAR-2000; 2000US-0190076.		
XX	18-APR-2000; 2000US-0198123.		
XX	19-MAY-2000; 2000US-0205515.		
XX	07-JUN-2000; 2000US-0209467.		
XX	28-JUN-2000; 2000US-0214886.		
XX	30-JUN-2000; 2000US-0215135.		
XX	07-JUL-2000; 2000US-0215647.		
XX	07-JUL-2000; 2000US-0216880.		
XX	11-JUL-2000; 2000US-0217487.		
XX	11-JUL-2000; 2000US-0217496.		
XX	14-JUL-2000; 2000US-0218290.		
XX	26-JUL-2000; 2000US-0220963.		
XX	26-JUL-2000; 2000US-0220964.		
XX	14-AUG-2000; 2000US-0224518.		
XX	14-AUG-2000; 2000US-0224519.		
XX	14-AUG-2000; 2000US-0225213.		
XX	14-AUG-2000; 2000US-0225214.		
XX	14-AUG-2000; 2000US-0225266.		
XX	14-AUG-2000; 2000US-0225267.		
XX	14-AUG-2000; 2000US-0225268.		
XX	14-AUG-2000; 2000US-0225270.		
XX	14-AUG-2000; 2000US-0225275.		
XX	14-AUG-2000; 2000US-0225758.		
XX	14-AUG-2000; 2000US-0225759.		
XX	18-AUG-2000; 2000US-0226279.		
XX	22-AUG-2000; 2000US-0226681.		
XX	22-AUG-2000; 2000US-0226686.		
XX	22-AUG-2000; 2000US-0227182.		
XX	23-AUG-2000; 2000US-0227009.		
XX	30-AUG-2000; 2000US-0228924.		
XX	01-SEP-2000; 2000US-0229287.		
XX	01-SEP-2000; 2000US-0229343.		
XX	01-SEP-2000; 2000US-0229344.		
XX	01-SEP-2000; 2000US-0229345.		
XX	05-SEP-2000; 2000US-0229509.		
XX	05-SEP-2000; 2000US-0229513.		
XX	06-SEP-2000; 2000US-0230437.		
XX	06-SEP-2000; 2000US-0230438.		
XX	08-SEP-2000; 2000US-0231242.		
XX	08-SEP-2000; 2000US-0231243.		
XX	08-SEP-2000; 2000US-0231244.		
XX	08-SEP-2000; 2000US-0231413.		
XX	08-SEP-2000; 2000US-0231414.		
XX	08-SEP-2000; 2000US-0232080.		
PR	08-SEP-2000; 2000US-0232081.		
PR	12-SEP-2000; 2000US-0231968.		
PR	14-SEP-2000; 2000US-0232397.		
PR	14-SEP-2000; 2000US-0232398.		
PR	14-SEP-2000; 2000US-0232399.		
PR	14-SEP-2000; 2000US-0232400.		
PR	14-SEP-2000; 2000US-0232401.		
PR	14-SEP-2000; 2000US-0233063.		
PR	14-SEP-2000; 2000US-0233064.		
PR	14-SEP-2000; 2000US-0233065.		
PR	21-SEP-2000; 2000US-0234223.		
PR	21-SEP-2000; 2000US-0234274.		
PR	25-SEP-2000; 2000US-0234997.		
PR	25-SEP-2000; 2000US-0234998.		
PR	25-SEP-2000; 2000US-0235484.		
PR	26-SEP-2000; 2000US-0235834.		
PR	27-SEP-2000; 2000US-0235836.		
PR	27-SEP-2000; 2000US-0236327.		
PR	29-SEP-2000; 2000US-0236367.		
PR	29-SEP-2000; 2000US-0236368.		
PR	29-SEP-2000; 2000US-0236369.		
PR	29-SEP-2000; 2000US-0236370.		
PR	29-SEP-2000; 2000US-0236802.		
PR	02-OCT-2000; 2000US-0237037.		
PR	02-OCT-2000; 2000US-0237038.		
PR	02-OCT-2000; 2000US-0237039.		
PR	02-OCT-2000; 2000US-0237040.		
PR	13-OCT-2000; 2000US-0239935.		
PR	13-OCT-2000; 2000US-0239937.		
PR	20-OCT-2000; 2000US-0240960.		
PR	20-OCT-2000; 2000US-0241221.		
PR	20-OCT-2000; 2000US-0241785.		
PR	20-OCT-2000; 2000US-0241786.		
PR	20-OCT-2000; 2000US-0241787.		
PR	20-OCT-2000; 2000US-0241808.		
PR	20-OCT-2000; 2000US-0241809.		
PR	20-OCT-2000; 2000US-0241826.		
PR	01-NOV-2000; 2000US-0244617.		
PR	08-NOV-2000; 2000US-0246474.		
PR	08-NOV-2000; 2000US-0246475.		
PR	08-NOV-2000; 2000US-0246476.		
PR	08-NOV-2000; 2000US-0246477.		
PR	08-NOV-2000; 2000US-0246478.		
PR	08-NOV-2000; 2000US-0246523.		
PR	08-NOV-2000; 2000US-0246524.		
PR	08-NOV-2000; 2000US-0246525.		
PR	08-NOV-2000; 2000US-0246526.		
PR	08-NOV-2000; 2000US-0246527.		
PR	08-NOV-2000; 2000US-0246528.		
PR	08-NOV-2000; 2000US-0246532.		
PR	08-NOV-2000; 2000US-0246609.		
PR	08-NOV-2000; 2000US-0246610.		
PR	08-NOV-2000; 2000US-0246611.		
PR	08-NOV-2000; 2000US-0246613.		
PR	17-NOV-2000; 2000US-0249207.		
PR	17-NOV-2000; 2000US-0249208.		
PR	17-NOV-2000; 2000US-0249209.		
PR	17-NOV-2000; 2000US-0249210.		
PR	17-NOV-2000; 2000US-0249211.		
PR	17-NOV-2000; 2000US-0249212.		
PR	17-NOV-2000; 2000US-0249213.		
PR	17-NOV-2000; 2000US-0249214.		
PR	17-NOV-2000; 2000US-0249215.		
PR	17-NOV-2000; 2000US-0249216.		
PR	17-NOV-2000; 2000US-0249217.		
PR	17-NOV-2000; 2000US-0249218.		
PR	17-NOV-2000; 2000US-0249244.		
PR	17-NOV-2000; 2000US-0249245.		
PR	17-NOV-2000; 2000US-0249264.		
PR	17-NOV-2000; 2000US-0249265.		
PR	17-NOV-2000; 2000US-0249297.		
PR	17-NOV-2000; 2000US-0249299.		
PR	17-NOV-2000; 2000US-0249300.		



CC treat bone marrow, blood and hematopoietic associated diseases and  
CC disorders, atelectasis, pulmonary congestion or oedema, emphysema,  
CC chronic bronchitis, bronchial asthma and bronchiectasis, intestinal  
CC disorders, spleen associated diseases, modulate renal disorders, treat  
CC cardiovascular disorders such as ischemic heart disease, modulate the  
CC proliferation, differentiation, and/or function of bone and cartilage  
CC cells and to treat bone and/or cartilage associated diseases or  
CC disorder. They may also be used to treat disorders associated with the  
CC ovaries, cerebral oedema, hydrocephalus, brain herniations, iatrogenic  
CC disease, inflammation, bacterial and viral meningitis, Alzheimer's  
CC disease, cerebral toxoplasmosis, Parkinson's disease, multiple sclerosis,  
CC brain cancers, hydrocephalus and encephalitis, and treat hepatic  
CC disorders.  
XX  
SQ Sequence 487 AA;

Query Match 50.1%; Score 770; DB 21; Length 487;  
Best Local Similarity 52.8%; Pred. No. 4.1e-74;  
Matches 150; Conservative 52; Mismatches 80; Indels 2; Gaps 1;

QY 11 PACYGGFDLYFLDKSGSVLHWNHNEIYFVEQLAHFISPOLRMSFIVSTRTTLMKLT 70  
Db 37 pscckafrdlyfvldksgsvannwiefvqhqrterfvspemlrslfivssqatillpt 96  
QY 71 EDREQIRQGLEELQKVLPGGDTYMHGFEFASQIYYENRQGYRTASVIITLTDGELHED 130  
Db 97 gdrykigkgledlkavkpvgetyihgklaneqi--qnagggkassiiialtdgkl dgl 154  
QY 131 LFFYSERANRDLGAIVYCVGVKDFNETQLARIADSKDHVPVNDGFOALOGIHSIL 190  
Db 155 vpsyaeneakrsigsvyvcgvldfeqaqleriadskdqvpkggfgalkginsil 214  
QY 191 KKSCITELAAPSTICAGESFQVVVRNGFRHARNVDRVLCSPKINDSVTLNEKPFSSVED 250  
Db 215 aqsceteilelspssvcvgekfqvltgravtsishdsgsvlctftanstytksekpvsiq 274  
QY 251 TYLLCPAPILKEVGMKAALQVMDGLSFSSSVIITTHGCSOG 294  
Db 275 ssilcpapvlnkdgetlevisyndgksavsrsltitatectng 318

Mus sp.  
WO200052022-A1.  
08-SEP-2000.  
01-MAR-2000; 2000WO-US05226.  
01-MAR-1999; 99US-0122458.

(MILL-) MILLENNIUM PHARM INC.  
Barnes TM, Holtzman DA, Sharp JD, Fraser CC;  
WPI; 2000-579269/54.  
N-PSDB; AAA75160.  
Novel human and murine secreted proteins designated TANGO 216, 261,  
262, 266 and 267 useful as modulating agents of cellular processes,  
e.g. for treating cancer -  
Disclosure; Page -: 175pp; English.  
AAB18458-60 represent murine TANGO 216 proteins. The specification also  
describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO  
polypeptides can be used to modulate cellular proliferation, modulate  
cellular differentiation and/or modulate cellular adhesion. The  
proteins can be used to treat any von Willebrand factor-associated  
disorder, regulate extracellular matrix structuring, cellular adhesion,  
and cell trafficking and/or migration, modulate cellular interactions,  
modulate cell adhesion in proliferative disorders, such as cancer,  
modulate the proliferation, differentiation, and/or function of cells  
that appear in the bone marrow, and leukocytes, treat bone marrow, blood  
and hematopoietic associated diseases and disorders, atelectasis,  
asthma and bronchiectasis, intestinal disorders, spleen associated  
diseases, modulate renal disorders, treat cardiovascular disorders such  
as ischemic heart disease, modulate the proliferation, differentiation,  
and/or function of bone and cartilage cells and to treat bone and/or  
cartilage associated diseases or disorder. They may also be used to  
treat disorders associated with the ovaries, and cerebral oedema,  
hydrocephalus, brain herniations, iatrogenic disease, inflammation,  
bacterial and viral meningitis, Alzheimer's Disease, cerebral  
toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers,  
hydrocephalus and encephalitis, and treat hepatic disorders.  
note: the present sequence does not appear in the specification; it was  
created using information provided.  
XX Sequence 487 AA;  
SQ

Query Match 50.1%; Score 770; DB 21; Length 487;  
Best Local Similarity 52.8%; Pred. No. 4.1e-74;  
Matches 150; Conservative 52; Mismatches 80; Indels 2; Gaps 1;

QY 11 PACYGGFDLYFLDKSGSVLHWNHNEIYFVEQLAHFISPOLRMSFIVSTRTTLMKLT 70  
Db 37 pscckafrdlyfvldksgsvannwiefvqhqrterfvspemlrslfivssqatillpt 96  
QY 71 EDREQIRQGLEELQKVLPGGDTYMHGFEFASQIYYENRQGYRTASVIITLTDGELHED 130  
Db 97 gdrykigkgledlkavkpvgetyihgklaneqi--qnagggkassiiialtdgkl dgl 154  
QY 131 LFFYSERANRDLGAIVYCVGVKDFNETQLARIADSKDHVPVNDGFOALOGIHSIL 190  
Db 155 vpsyaeneakrsigsvyvcgvldfeqaqleriadskdqvpkggfgalkginsil 214  
QY 191 KKSCITELAAPSTICAGESFQVVVRNGFRHARNVDRVLCSPKINDSVTLNEKPFSSVED 250  
Db 215 aqsceteilelspssvcvgekfqvltgravtsishdsgsvlctftanstytksekpvsiq 274  
QY 251 TYLLCPAPILKEVGMKAALQVMDGLSFSSSVIITTHGCSOG 294  
Db 275 ssilcpapvlnkdgetlevisyndgksavsrsltitatectng 318

RESULT 13  
AAB18460  
ID AAB18460 standard; Protein; 487 AA.  
XX  
AC AAB18460;  
XX  
DT 15-JAN-2001 (first entry)

XX DE A murine TANGO 216 polypeptide clone.

XX KW TANGO 266; TANGO 216;; TANGO 261; TANGO 262; TANGO 267;

KW cellular proliferation; cellular differentiation; cellular adhesion;

KW von Willebrand factor-associated disorder; cell trafficking; cancer;

KW hematopoietic associated disease; atelectasis; pulmonary congestion;

KW oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;

KW intestinal disorder; spleen associated disease; renal disorder;

KW cardiovascular disorder; ischemic heart disease; hydrocephalus;

KW brain herniation; iatrogenic disease; inflammation; meningitis;

KW Alzheimer's Disease; cerebral toxoplasmosis; Parkinson's disease;

KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.

XX OS Mus sp.

XX PN WO200052022-A1.

XX PD 08-SEP-2000.

XX PF 01-MAR-2000; 2000WO-US05226.

XX PR 01-MAR-1999; 99US-0122458.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Barnes TM, Holtzman DA, Sharp JD, Fraser CC;

XX KW WPI; 2000-579269/54.

DR N-PSDB; AAA75162.

XX PT Novel human and murine secreted proteins designated TANGO 216, 261,

PT 262, 266 and 267 useful as modulating agents of cellular processes,

PT e.g. for treating cancer -

PS Disclosure; Page -: 175pp; English.

XX KW AAB18458-60 represent murine TANGO 216 proteins. The specification also

CC describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO

CC polypeptides can be used to modulate cellular proliferation, modulate

CC cellular differentiation and/or modulate cellular adhesion. The

CC proteins can be used to treat any von Willebrand factor-associated

CC disorder, regulate extracellular matrix structuring, cellular adhesion,

CC and cell trafficking and/or migration, modulate cellular interactions,

CC modulate cell adhesion in proliferative disorders, such as cancer,

CC modulate the proliferation, differentiation, and/or function of cells

CC that appear in the bone marrow, and leukocytes, treat bone marrow, blood

CC and hematopoietic associated diseases and disorders, atelectasis,

CC pulmonary congestion or oedema, emphysema, chronic bronchitis, bronchial

CC asthma and bronchiectasis, intestinal disorders, spleen associated

CC diseases, modulate renal disorders, treat cardiovascular disorders such

CC as ischemic heart disease, modulate the proliferation, differentiation,

CC and/or function of bone and cartilage cells and to treat bone and/or

CC cartilage associated diseases or disorder. They may also be used to

CC treat disorders associated with the ovaries, and cerebral oedema,

CC hydrocephalus, brain herniations, iatrogenic disease, inflammations,

CC bacterial and viral meningitis, Alzheimer's disease, cerebral

CC toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers,

CC hydrocephalus and encephalitis, and treat hepatic disorders.

CC note: the present sequence does not appear in the specification; it was

CC created using information provided.

XX Sequence 487 AA;

Query Match 50.1%; Score 769; DB 21; Length 487;

Best Local Similarity 52.8%; Pred. No. 5.3e-74;

Matches 150; Conservative 52; Mismatches 80; Indels 2; Gaps 1;

QY 11 PACYGGFDLYLDKSGVLHWHNEIYFVEQLAHKFISPOLRMSFIVFSTRGTYLMLT 70

Db 37 psckkafdlrvldksgvannwieiynfvhqltdrfvpspmrisflvfssqatliplt 96

QY 71 EDREIQROGLEELQKVLPGGDTYMHGFERASEQIYYENRGYRTASVIALTDCELHED 130

Db 97 gdrykigkgledlkavkpvgetyihgklaneql--qnagglkassilaltgkidgl 154

QY 131 LFFYSERANRSDLGAIVYCVGVKDFNETQLARIADSKDHVPVNDGFMALQGIHSL 190

Db 155 vpsyaeneakksrslgasvycvgvldfeqaqlerladskdqvfvpkggfgalkglnsil 214

QY 191 KKSCEITLAAEPSTICAGESFOVVVRGNGFRHARNVDRVLCSPKINDSVTLNEKPFVED 250

Db 215 aqscstellespssvcvgvgekfqvvtgravtslshdgsvicftfctanscytksekpslqp 274

QY 251 TYLCPAPILKEVGMKAALQVMMNDGLSFSSSVIITTHCSGDG 294

Db 275 ssilcpapvinkdgetlievsisyndgksavrsititactctng 318

RESULT 14

AAB18459

ID AAB18459 standard; Protein; 487 AA.

XX AC AAB18459;

XX DT 15-JAN-2001 (first entry)

XX DE A murine TANGO 216 polypeptide clone.

XX KW TANGO 266; TANGO 216;; TANGO 261; TANGO 262; TANGO 267;

KW cellular proliferation; cellular differentiation; cellular adhesion;

KW von Willebrand factor-associated disorder; cell trafficking; cancer;

KW hematopoietic associated disease; atelectasis; pulmonary congestion;

KW oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;

KW intestinal disorder; spleen associated disease; renal disorder;

KW cardiovascular disorder; ischemic heart disease; hydrocephalus;

KW brain herniation; iatrogenic disease; inflammation; meningitis;

KW Alzheimer's Disease; cerebral toxoplasmosis; Parkinson's disease;

KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder.

XX OS Mus sp.

XX PN WO200052022-A1.

XX PD 08-SEP-2000.

XX PF 01-MAR-2000; 2000WO-US05226.

XX PR 01-MAR-1999; 99US-0122458.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Barnes TM, Holtzman DA, Sharp JD, Fraser CC;

XX KW WPI; 2000-579269/54.

DR N-PSDB; AAA75161.

XX PT Novel human and murine secreted proteins designated TANGO 216, 261,

PT 262, 266 and 267 useful as modulating agents of cellular processes,

PT e.g. for treating cancer -

PS Disclosure; Page -: 175pp; English.

XX KW AAB18458-60 represent murine TANGO 216 proteins. The specification also

CC describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO

CC polypeptides can be used to modulate cellular proliferation, modulate

CC cellular differentiation and/or modulate cellular adhesion. The

CC proteins can be used to treat any von Willebrand factor-associated

CC disorder, regulate extracellular matrix structuring, cellular adhesion,

CC and cell trafficking and/or migration, modulate cellular interactions,

CC modulate cell adhesion in proliferative disorders, such as cancer,

CC modulate the proliferation, differentiation, and/or function of cells

CC that appear in the bone marrow, and leukocytes, treat bone marrow, blood

CC and hematopoietic associated diseases and disorders, atelectasis,

CC pulmonary congestion or oedema, emphysema, chronic bronchitis, bronchial

CC asthma and bronchiectasis, intestinal disorders, spleen associated

CC diseases, modulate renal disorders, treat cardiovascular disorders such

CC as ischemic heart disease, modulate the proliferation, differentiation,

CC and/or function of bone and cartilage cells and to treat bone and/or

CC cartilage associated diseases or disorder. They may also be used to

CC treat disorders associated with the ovaries, and cerebral oedema,

CC hydrocephalus, brain herniations, iatrogenic disease, inflammations,

CC bacterial and viral meningitis, Alzheimer's disease, cerebral

CC toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers,

CC hydrocephalus and encephalitis, and treat hepatic disorders.

CC note: the present sequence does not appear in the specification; it was

CC created using information provided.

Fri Aug 9 10:57:01 2002

CC asthma and bronchiectasis, intestinal disorders, spleen associated  
CC diseases, modulate renal disorders, treat cardiovascular disorders such  
CC as ischemic heart disease, modulate the proliferation, differentiation,  
CC and/or function of bone and cartilage cells and to treat bone and/or  
CC cartilage associated diseases or disorder. They may also be used to  
CC treat disorders associated with the ovaries, and cerebral oedema,  
CC hydrocephalus, brain herniations, iatrogenic disease, inflammations,  
CC bacterial and viral meningitis, Alzheimer's Disease, cerebral  
CC toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers,  
CC hydrocephalus and encephalitis, and treat hepatic disorders.  
CC note: the present sequence does not appear in the specification; it was  
CC created using information provided.

XX Sequence 487 AA;  
SQ  
Query Match 49.9%; Score 767; DB 21; Length 487;  
Best Local Similarity 52.5%; Pred. No. 8.7e-74;  
Matches 149; Conservative 53; Mismatches 80; Indels 2; Gaps 1;  
QY 11 PACYGGFDLYFLDKSGVLHWHNEIYFVEQLAKHFISPLRMSFIVFSTRTMLKLT 70  
DB 37 pscckafolyfvlcksgsvanwidynfvhqlterfvpemrlsfivfssqatiliplt 96  
QY 71 EDREIQRLGLELQKVLPGDYYWHEGFERASEQIYYENRGYFTASVIITLDTGELHED 130  
DB 97 gdrykigkgledlkavkpvetihgiklaneqi--qnagglkassilialtdgkldgl 154  
QY 131 LEFFYSERANRSRDLAGVYCVGVKDFNETQLARIADSKDHVPFVNDGFOALOGIIHSIL 190  
DB 155 vpsyaenackrsrgasvycvgvldfcaqleriadskdqvfpkggfkalkginsil 214  
QY 191 KKSCIEIIAAEPSTICAGESFQVVRNGFRHARNVDRVLCFSKINDSVTLNEKPFVSD 250  
DB 215 agscleilelpsvvcgkfqvltgravtsishdsgvltctfstanstytcksekvsiqp 274  
QY 251 TYLLCPAPILKEVGMKAALQVSMNDGLSFSSVITTTTHGSDG 294  
DB 275 ssilcpapvinkdgetlevisyngdksavrsrltitatectng 318

RESULT 15  
AAB01428  
ID AAB01428 standard; Protein; 381 AA.  
XX  
AC AAB01428;  
XX  
DT 20-OCT-2000 (first entry)  
XX  
DE Murine TANGO 197.  
XX  
KW TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;  
KW graft versus-host diseases; rheumatoid arthritis; psoriasis;  
KW inflammatory bowel disease; septic shock; ulcerative colitis;  
KW Crohn's disease; chronic myelogenous leukemia; cancer; liver  
KW disease; Hodgkin's disease; osteoarthritis; Lyme's disease;  
KW cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;  
KW systemic lupus erythematosus; transgenic animal; diagnosis;  
KW prognosis; prophylactic; therapeutic; mouse.  
XX  
OS Mus musculus.  
XX  
PN WO200039284-A1.  
XX  
PD 06-JUL-2000.  
XX  
XX 23-DEC-1999; 99WO-US31025.  
XX  
PR 30-DEC-1998; 98US-02233546.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Holtzman DA;

XX WPI; 2000-465743/40.  
DR N-PSDB; AAA47479.  
XX  
XX Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,  
PT 224 and 239 polypeptides useful for the treatment of asthma, rheumatoid  
PT arthritis, psoriasis and autoimmune diseases  
XX  
XX Claim 8; Fig 27; 209pp; English.  
XX  
XX Nucleic acids encoding TANGO polypeptides are useful as modulating  
CC agents for regulating cellular processes like asthma, graft  
CC versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory  
CC bowel disease, septic shock, ulcerative colitis, Crohn's disease,  
CC chronic myelogenous leukemia, cancer, liver disease, Hodgkin's  
CC disease, osteoarthritis, Lyme's disease, cachexia and autoimmune  
CC diseases e.g. myasthenia gravis, autoimmune diabetes and systemic  
CC lupus erythematosus. The nucleic acids are also useful for producing  
CC transgenic animals and the TANGO polypeptides themselves. Partial  
CC TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in  
CC forensic biology, for diagnostic assays, prognostic assays,  
CC pharmacogenomics and for monitoring clinical trials. TANGO  
CC polypeptides are suitable for both prophylactic and therapeutic  
CC methods for treating a subject at risk of a disorder or having a  
CC disorder associated with aberrant TANGO expression. A wide range  
CC of cellular disorders can be treated.  
XX  
SQ Sequence 381 AA;

Query Match 44.6%; Score 685; DB 21; Length 381;  
Best Local Similarity 99.3%; Pred. No. 4.3e-65;  
Matches 134; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 161 QLARIADSKDHVPFVNDGFOALOGIIHSILKSCIEIIAAEPSTICAGESFQVVRNGF 220  
DB 4 qlariadskdhvfpvndgfqalqglihskscieilaepsticagesfqvvrngf 63  
QY 221 RHARNVDRVLCFSKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMNDGLSFI 280  
DB 64 rharnvdrvicskindsvtlnekpavedtyllcpapilkevgmkaalqvsmndglsfi 123  
QY 281 SSSVITTTTHGSDGS 295  
DB 124 sssvltttthcsdgs 138

Search completed: August 9, 2002, 10:32:11  
Job time: 130 sec



---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:30:02 ; Search time 98.32 Seconds  
(without alignments)  
656.123 Million cell updates/sec

Title: US-09-970-076-2

Perfect score: 1914

Sequence: 1 MATAERRALIGFQWLSLAT.....VTKVPPPPAESEENKIK 368

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 528882 seqs, 17529045 residues

Total number of hits satisfying chosen parameters: 528882

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_New.\*

- 1: /cgn2.6/ptodata/2/paa/PCT\_NEW\_COMB.pcp.\*
- 2: /cgn2.6/ptodata/2/paa/US06\_NEW\_COMB.pcp.\*
- 3: /cgn2.6/ptodata/2/paa/US07\_NEW\_COMB.pcp.\*
- 4: /cgn2.6/ptodata/2/paa/US08\_NEW\_COMB.pcp.\*
- 5: /cgn2.6/ptodata/2/paa/US09\_NEW\_COMB.pcp.\*
- 6: /cgn2.6/ptodata/2/paa/US10\_NEW\_COMB.pcp.\*
- 7: /cgn2.6/ptodata/2/paa/US60\_NEW\_COMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1894	99.0	564	1	PCT-US02-08253-187
2	1894	99.0	564	1	PCT-US02-08253-232
3	1894	99.0	564	7	US-60-389-987-1823
4	1793	93.7	562	1	PCT-US02-08253-194
5	1793	93.7	562	1	PCT-US02-08253-301
6	1661	86.8	551	6	US-10-038-307-18
7	1849	86.2	333	6	US-10-038-307-2
8	1649	86.2	345	6	US-10-038-307-24
9	1649	86.2	564	6	US-10-038-307-20
10	1636	85.5	328	6	US-10-038-307-26
11	1634.5	85.4	342	6	US-10-038-307-22
12	1629	85.1	543	6	US-10-038-307-14
13	1629	85.1	543	6	US-10-038-307-16
14	1619	84.6	543	6	US-10-038-307-10
15	1488	77.7	534	6	US-10-038-307-12
16	966	50.5	218	5	US-09-629-469A-11706
17	962.5	50.3	488	6	US-10-104-047-2639
18	957.5	50.0	488	6	US-10-038-307-6
19	957.5	50.0	488	7	US-60-373-595-10
20	938.5	49.0	587	6	US-10-125-540-312
21	936.5	48.9	487	6	US-10-038-307-8
22	909	47.5	381	6	US-10-038-307-4
23	903	47.2	538	6	US-10-047-542-99
24	145.5	7.6	1152	5	US-09-592-617A-43
25	138.5	7.2	1163	1	PCT-US02-10824-116
26	138.5	7.2	1163	7	US-60-369-452-33

27 134.5 7.0 646 6 US-10-155-881-22380 Sequence 22380, A  
28 132.5 6.9 191 5 US-09-805-354-1 Sequence 1, Appl  
29 132 6.9 460 6 US-10-155-881-33525 Sequence 33525, A  
30 131 6.8 3063 1 PCT-US02-08253-257 Sequence 257, App  
31 131 6.8 3063 1 PCT-US02-19669-61 Sequence 61, Appl  
32 131 6.8 3063 1 PCT-US02-19669-63 Sequence 63, Appl  
33 131 6.8 3063 6 US-10-177-293-61 Sequence 61, Appl  
34 131 6.8 3063 6 US-10-177-293-63 Sequence 63, Appl  
35 128 6.7 3051 7 US-60-360-039-5866 Sequence 5866, Ap  
36 125.5 6.6 191 5 US-09-805-354-2 Sequence 2, Appl  
37 124.5 6.5 191 5 US-09-805-354-3 Sequence 3, Appl  
38 121.5 6.3 187 5 US-09-592-617A-59 Sequence 59, Appl  
39 118.5 6.2 3594 6 US-10-150-821-4 Sequence 4, Appl  
40 117.5 6.1 1207 7 US-60-389-987-1591 Sequence 1591, Ap  
41 115 6.0 682 6 US-10-155-881-28757 Sequence 28757, A  
42 113.5 5.9 764 1 PCT-US02-19669-27 Sequence 27, Appl  
43 113.5 5.9 764 6 US-10-177-293-27 Sequence 27, Appl  
44 113.5 5.9 795 6 US-10-009-557-7 Sequence 7, Appl  
45 113.5 5.9 915 5 US-09-904-920A-34 Sequence 34, Appl

#### ALIGNMENTS

RESULT 1

PCT-US02-08253-187

; Sequence 187, Application PC/TUS0208253

; GENERAL INFORMATION:

; APPLICANT: Carson-Walter, Eleanor

; APPLICANT: St. Croix, Brad

; APPLICANT: Vogelstein, Bert

; APPLICANT: Kinzler, Kenneth

; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS

; FILE REFERENCE: 1107.00179

; CURRENT APPLICATION NUMBER: PCT/US02/08253

; PRIOR FILING DATE: 2002-04-10

; PRIOR APPLICATION NUMBER: 60/282,850

; PRIOR FILING DATE: 2001-04-11

; PRIOR APPLICATION NUMBER: 60/308,829

; PRIOR FILING DATE: 2001-08-01

; NUMBER OF SEQ ID NOS: 359

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 187

; LENGTH: 564

; TYPE: PRT

; ORGANISM: Homo sapiens

PCT-US02-08253-187

Query Match 99.0%; Score 1894; DB 1; Length 564;  
Best Local Similarity 100.0%; Pred. No. 2.1e-166;  
Matches 364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATAERRALIGFQWLSLATLVLCAGQGRREDGGPACYGFDLYLDKSGSVLHHWN 60  
DB 1 MATAERRALIGFQWLSLATLVLCAGQGRREDGGPACYGFDLYLDKSGSVLHHWN 60  
QY 61 EIIYFVEQLAHKFISPOLRMSFIVSTRTTLMKLTEDREQIROLGLEELQKVLPGGDTYM 120  
DB 61 EIIYFVEQLAHKFISPOLRMSFIVSTRTTLMKLTEDREQIROLGLEELQKVLPGGDTYM 120  
QY 121 HEGFERASQIYYENRQGYRTASVIITADTDELHEDLFFYSERANRSRDLGAIYVCVG 180  
DB 121 HEGFERASQIYYENRQGYRTASVIITADTDELHEDLFFYSERANRSRDLGAIYVCVG 180  
QY 181 KDFNETQLARIADSKDHVPVNDGFOALOGIITHSLKSCIEILAAEPSTICAGESFQVW 240  
DB 181 KDFNETQLARIADSKDHVPVNDGFOALOGIITHSLKSCIEILAAEPSTICAGESFQVW 240  
QY 241 VRNGFRHARNVDRLVLCSPKINDSVTLNKPFSVEDTYLLCPAPILKEVGMKAALQVSMN 300  
DB 241 VRNGFRHARNVDRLVLCSPKINDSVTLNKPFSVEDTYLLCPAPILKEVGMKAALQVSMN 300

Fri Aug 9 10:57:00 2002

```

; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465p2
; CURRENT APPLICATION NUMBER: US/60/389,987
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 3025
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1823
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-60-389-987-1823

Query Match          99.0%; Score 1894; DB 7; Length 564;
Best Local Similarity 100.0%; Pred. No. 2.le-166;
Matches 364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATAERRALGIGFOWLSLATLVLCACGGRRDGGPACYGFDLYFILDKSGSVLHHWN 60
Db 1 MATAERRALGIGFOWLSLATLVLCACGGRRDGGPACYGFDLYFILDKSGSVLHHWN 60
QY 61 EIIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDVTM 120
Db 61 EIIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDVTM 120
QY 121 HEGFERASEQIYYENRQGYRTASVIITDGTGELHEDLFFYSEREANRSDLGAIYVCVG 180
Db 121 HEGFERASEQIYYENRQGYRTASVIITDGTGELHEDLFFYSEREANRSDLGAIYVCVG 180
QY 181 KDFNETOLARIADSKDHVPVNDGFGQALQGIHSLKSCIEILAAEPSTICAGESFQV 240
Db 181 KDFNETOLARIADSKDHVPVNDGFGQALQGIHSLKSCIEILAAEPSTICAGESFQV 240
QY 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300
Db 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300
QY 301 DGLSFSSSVIITTHCSGSIILAIALLILFLALLLALMLWFWPLCCTVIKEVPPPPAE 360
Db 301 DGLSFSSSVIITTHCSGSIILAIALLILFLALLLALMLWFWPLCCTVIKEVPPPPAE 360
QY 361 ESEE 364
Db 361 ESEE 364

RESULT 4
PCT-US02-08253-194
; Sequence 194, Application PC/TUS0208253
; GENERAL INFORMATION:
; APPLICANT: Carson-Walter, Eleanor
; APPLICANT: St. Croix, Brad
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00179
; CURRENT APPLICATION NUMBER: PCT/US02/08253
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/308,829
; PRIOR FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 359
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 194
; LENGTH: 562
; TYPE: PRT

; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465p2
; CURRENT APPLICATION NUMBER: US/60/389,987
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 3025
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1823
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-60-389-987-1823

Query Match          99.0%; Score 1894; DB 1; Length 564;
Best Local Similarity 100.0%; Pred. No. 2.le-166;
Matches 364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATAERRALGIGFOWLSLATLVLCACGGRRDGGPACYGFDLYFILDKSGSVLHHWN 60
Db 1 MATAERRALGIGFOWLSLATLVLCACGGRRDGGPACYGFDLYFILDKSGSVLHHWN 60
QY 61 EIIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDVTM 120
Db 61 EIIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDVTM 120
QY 121 HEGFERASEQIYYENRQGYRTASVIITDGTGELHEDLFFYSEREANRSDLGAIYVCVG 180
Db 121 HEGFERASEQIYYENRQGYRTASVIITDGTGELHEDLFFYSEREANRSDLGAIYVCVG 180
QY 181 KDFNETOLARIADSKDHVPVNDGFGQALQGIHSLKSCIEILAAEPSTICAGESFQV 240
Db 181 KDFNETOLARIADSKDHVPVNDGFGQALQGIHSLKSCIEILAAEPSTICAGESFQV 240
QY 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300
Db 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFVSVEDTYLLCPAPILKEVGMKAALQVSMN 300
QY 301 DGLSFSSSVIITTHCSGSIILAIALLILFLALLLALMLWFWPLCCTVIKEVPPPPAE 360
Db 301 DGLSFSSSVIITTHCSGSIILAIALLILFLALLLALMLWFWPLCCTVIKEVPPPPAE 360
QY 361 ESEE 364
Db 361 ESEE 364

RESULT 3
US-60-389-987-1823
; Sequence 1823, Application US/60389987
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Faby, Eoin D.
```

```
; ORGANISM: Mus musculus
PCT-US02-08253-194

Query Match      93.7%; Score 1793; DB 1; Length 562;
Best Local Similarity 96.1%; Pred. No. 4.4e-157;
Matches 342; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 9 LGIGFQWLSLATLVLCAGGRRDGGPACYGDFDLYFILDKSGSVLHHNEIYFVEQ 68
Db 7 LGAGLUGLCVAALVLCAGHGRDGGPACYGDFDLYFILDKSGSVLHHNEIYFVEQ 66
Qy 69 LAHKFISPOLRMSFVSTRTTLMKLTEDREQIROGLEELQKVLPGGDTYMHGFERAS 128
Db 67 LAHREISPOLRMSFVSTRTTLMKLTEDREQIROGLEELQKVLPGGDTYMHGFERAS 126
Qy 129 EQIYYENRQGYRTASVIITDGLHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQL 188
Db 127 EQIYYENRQGYRTASVIITDGLHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQL 186
Qy 189 ARIADSKDHVPVNDGFOALOGIIHSILKSCIEILAAEPSTICAGESFOVVVRNGFRH 248
Db 187 ARIADSKDHVPVNDGFOALOGIIHSILKSCIEILAAEPSTICAGESFOVVVRNGFRH 246
Qy 249 ARNDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMNDGLSFSS 308
Db 247 ARNDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMNDGLSFSS 306
Qy 309 SVITTTCHSCDGSILAIALLFLLLALALLWFWPLCCTVTKVPPPPAESEE 364
Db 307 SVIITTHCHSCDGSILAIALLVLLALLWFWPLCCTVTKVPPPPAESEE 362

RESULT 5
PCT-US02-08253-301
; Sequence 301, Application PC/TUS0208253
; GENERAL INFORMATION:
; APPLICANT: Carson-Walter, Eleanor
; APPLICANT: St. Croix, Brad
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00179
; CURRENT APPLICATION NUMBER: PCT/US02/08253
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/308,829
; PRIOR FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 359
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 301
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Mouse
PCT-US02-08253-301

Query Match      93.7%; Score 1793; DB 1; Length 562;
Best Local Similarity 96.1%; Pred. No. 4.4e-157;
Matches 342; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 9 LGIGFQWLSLATLVLCAGGRRDGGPACYGDFDLYFILDKSGSVLHHNEIYFVEQ 68
Db 7 LGAGLUGLCVAALVLCAGHGRDGGPACYGDFDLYFILDKSGSVLHHNEIYFVEQ 66
Qy 69 LAHKFISPOLRMSFVSTRTTLMKLTEDREQIROGLEELQKVLPGGDTYMHGFERAS 128
Db 67 LAHREISPOLRMSFVSTRTTLMKLTEDREQIROGLEELQKVLPGGDTYMHGFERAS 126
Qy 129 EQIYYENRQGYRTASVIITDGLHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQL 188
Db 127 EQIYYENRQGYRTASVIITDGLHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQL 186
```

```
; ORGANISM: Mus musculus
PCT-US02-08253-194

Query Match      93.7%; Score 1793; DB 1; Length 562;
Best Local Similarity 96.1%; Pred. No. 4.4e-157;
Matches 342; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 9 LGIGFQWLSLATLVLCAGGRRDGGPACYGDFDLYFILDKSGSVLHHNEIYFVEQ 68
Db 7 LGAGLUGLCVAALVLCAGHGRDGGPACYGDFDLYFILDKSGSVLHHNEIYFVEQ 66
Qy 69 LAHKFISPOLRMSFVSTRTTLMKLTEDREQIROGLEELQKVLPGGDTYMHGFERAS 128
Db 67 LAHREISPOLRMSFVSTRTTLMKLTEDREQIROGLEELQKVLPGGDTYMHGFERAS 126
Qy 129 EQIYYENRQGYRTASVIITDGLHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQL 188
Db 127 EQIYYENRQGYRTASVIITDGLHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQL 186

; ORGANISM: Homo sapiens
US-10-038-307-18

Query Match      86.8%; Score 1661; DB 6; Length 551;
Best Local Similarity 100.0%; Pred. No. 6.8e-145;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATAERRALGIGFQWLSLATLVLCAGGRRDGGPACYGDFDLYFILDKSGSVLHHN 60
Db 1 MATAERRALGIGFQWLSLATLVLCAGGRRDGGPACYGDFDLYFILDKSGSVLHHN 60
Qy 61 EIYFVEQLAHKFISPOLRMSFVSTRTTLMKLTEDREQIROGLEELQKVLPGGDTY 120
Db 61 EIYFVEQLAHKFISPOLRMSFVSTRTTLMKLTEDREQIROGLEELQKVLPGGDTY 120
Qy 121 HEGFERASEQIYYENRQGYRTASVIITDGLHEDLFFYSEREANRSDLGAIYVCV 180
Db 121 HEGFERASEQIYYENRQGYRTASVIITDGLHEDLFFYSEREANRSDLGAIYVCV 180
Qy 181 KDFNETQARIADSKDHVPVNDGFOALOGIIHSILKSCIEILAAEPSTICAGESFO 240
Db 181 KDFNETQARIADSKDHVPVNDGFOALOGIIHSILKSCIEILAAEPSTICAGESFO 240
Qy 241 VRGNGFRHARNVDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQ 300
Db 241 VRGNGFRHARNVDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQ 300
Qy 301 DGLSFSSSVIITTHCHSDG 320
Db 301 DGLSFSSSVIITTHCHSDG 320

RESULT 7
US-10-038-307-2
; Sequence 2, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OzKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
```

Fri Aug 9 10:57:00 2002

FILE REFERENCE: 7853-253-999  
CURRENT APPLICATION NUMBER: US/10/038, 307  
CURRENT FILING DATE: 2002-06-28  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 333  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-038-307-2

Query Match 86.2%; Score 1649; DB 6; Length 333;  
Best Local Similarity 100.0%; Pred. No. 4.4e-144;  
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATAERRALGIGFOWLSLTLVLCAGGGRRDGGPACYGFDLYFLDKSGSVLHHN 60  
Db 1 MATAERRALGIGFOWLSLTLVLCAGGGRRDGGPACYGFDLYFLDKSGSVLHHN 60  
Qy 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Db 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Qy 121 HEGFERASEQIYYENROGYRTASVITLTDGELHEDLFFYSEREANKSRDLGAIVYCVGV 180  
Db 121 HEGFERASEQIYYENROGYRTASVITLTDGELHEDLFFYSEREANKSRDLGAIVYCVGV 180  
Qy 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHSLKSKSCIEILAAEPSTICAGESFOV 240  
Db 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHSLKSKSCIEILAAEPSTICAGESFOV 240  
Qy 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Qy 301 DGLSFSSSVIIITTHCS 318  
Db 301 DGLSFSSSVIIITTHCS 318

RESULT 8  
US-10-038-307-24  
Sequence 24, Application US/10038307  
GENERAL INFORMATION:  
APPLICANT: James B. ROTTMAN  
APPLICANT: Theresa L. O'KEEFE  
APPLICANT: Engin OZKAYNAK  
APPLICANT: Judith J. HEALEY  
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
FILE REFERENCE: 7853-253-999  
CURRENT APPLICATION NUMBER: US/10/038, 307  
CURRENT FILING DATE: 2002-06-28  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 24  
LENGTH: 345  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-038-307-24

Query Match 86.2%; Score 1649; DB 6; Length 345;  
Best Local Similarity 100.0%; Pred. No. 4.6e-144;  
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATAERRALGIGFOWLSLTLVLCAGGGRRDGGPACYGFDLYFLDKSGSVLHHN 60  
Db 1 MATAERRALGIGFOWLSLTLVLCAGGGRRDGGPACYGFDLYFLDKSGSVLHHN 60  
Qy 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Db 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120

Qy 121 HEGFERASEQIYYENROGYRTASVITLTDGELHEDLFFYSEREANKSRDLGAIVYCVGV 180  
Db 121 HEGFERASEQIYYENROGYRTASVITLTDGELHEDLFFYSEREANKSRDLGAIVYCVGV 180  
Qy 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHSLKSKSCIEILAAEPSTICAGESFOV 240  
Db 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHSLKSKSCIEILAAEPSTICAGESFOV 240  
Qy 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Qy 301 DGLSFSSSVIIITTHCS 318  
Db 301 DGLSFSSSVIIITTHCS 318

RESULT 9  
US-10-038-307-20  
Sequence 20, Application US/10038307  
GENERAL INFORMATION:  
APPLICANT: James B. ROTTMAN  
APPLICANT: Theresa L. O'KEEFE  
APPLICANT: Engin OZKAYNAK  
APPLICANT: Judith J. HEALEY  
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
FILE REFERENCE: 7853-253-999  
CURRENT APPLICATION NUMBER: US/10/038, 307  
CURRENT FILING DATE: 2002-06-28  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 20  
LENGTH: 564  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-038-307-20

Query Match 86.2%; Score 1649; DB 6; Length 564;  
Best Local Similarity 100.0%; Pred. No. 9e-144;  
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATAERRALGIGFOWLSLTLVLCAGGGRRDGGPACYGFDLYFLDKSGSVLHHN 60  
Db 1 MATAERRALGIGFOWLSLTLVLCAGGGRRDGGPACYGFDLYFLDKSGSVLHHN 60  
Qy 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Db 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Qy 121 HEGFERASEQIYYENROGYRTASVITLTDGELHEDLFFYSEREANKSRDLGAIVYCVGV 180  
Db 121 HEGFERASEQIYYENROGYRTASVITLTDGELHEDLFFYSEREANKSRDLGAIVYCVGV 180  
Qy 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHSLKSKSCIEILAAEPSTICAGESFOV 240  
Db 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHSLKSKSCIEILAAEPSTICAGESFOV 240  
Qy 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 241 VRNGFRHARNVDRVLCFSKINDSVTLNEKPFSEVEDTYLLCPAPILKEVGMKAALQVSMN 300  
Qy 301 DGLSFSSSVIIITTHCS 318  
Db 301 DGLSFSSSVIIITTHCS 318

RESULT 10  
US-10-038-307-26  
Sequence 26, Application US/10038307  
GENERAL INFORMATION:  
APPLICANT: James B. ROTTMAN

```
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
us-10-038-307-26

Query Match      85.5%; Score 1636; DB 6; Length 328;
Best Local Similarity 100.0%; Pred. No. 6.e-143;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATAERRALGIGFQWLSLATLVLCAGOGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60
Db 1 MATAERRALGIGFQWLSLATLVLCAGOGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60
Qy 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120
Db 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120
Qy 121 HEGFERASEQIYYENRQGYRTASVITATDGHEDHDLFFYSEREANRSDLGAIYVCVG 180
Db 121 HEGFERASEQIYYENRQGYRTASVITATDGHEDHDLFFYSEREANRSDLGAIYVCVG 180
Qy 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHSLILKKSCEIILAAEPSTICAGESFQV 240
Db 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHSLILKKSCEIILAAEPSTICAGESFQV 240
Qy 241 VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300
Db 241 VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300
Qy 301 DGLSFSSSVIITTH 316
Db 301 DGLSFSSSVIITTH 316

RESULT 11
us-10-038-307-22
; Sequence 22, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Homo sapiens
us-10-038-307-22

Query Match      85.4%; Score 1634.5; DB 6; Length 342;
Best Local Similarity 97.2%; Pred. No. 9.9e-143;
Matches 318; Conservative 0; Mismatches 0; Indels 9; Gaps 1;

Qy 1 MATAERRALGIGFQWLSLATLVLCAGOGGRRDGGPACYGFDLYFLDKSGSVLHHWN 51
Db 1 MATAERRALGIGFQWLSLATLVLCAGOGGRRDGGPACYGFDLYFLDKSGSVLHHWN 51
```

```
Qy 52 SGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQK 111
Db 61 SGSVLHHWNEIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQK 120
Qy 112 VLPGGDTYMHGFERASEQIYYENRQGYRTASVITATDGHEDHDLFFYSEREANRSD 171
Db 121 VLPGGDTYMHGFERASEQIYYENRQGYRTASVITATDGHEDHDLFFYSEREANRSD 180
Qy 172 GAIVYCVGKDFNETQLARIADSKDHVPVNDGFOALQGIHSLILKKSCEIILAAEPSTI 231
Db 181 GAIVYCVGKDFNETQLARIADSKDHVPVNDGFOALQGIHSLILKKSCEIILAAEPSTI 240
Qy 232 CAGESFQVVRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGM 291
Db 241 CAGESFQVVRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGM 300
Qy 292 KAALQVSMNDGLSFSSSVIITTHCS 318
Db 301 KAALQVSMNDGLSFSSSVIITTHCS 327

RESULT 12
us-10-038-307-14
; Sequence 14, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Homo sapiens
us-10-038-307-14

Query Match      85.1%; Score 1629; DB 6; Length 543;
Best Local Similarity 98.8%; Pred. No. 6e-142;
Matches 317; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

Qy 1 MATAERRALGIGFQWLSLATLVLCAGOGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60
Db 1 MATAERRALGIGFQWLSLATLVLCAGOGGRRDGGPACYGFDLYFLDKSGSVLHHWN 60
Qy 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120
Db 61 EIIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGDITYM 120
Qy 121 HEGFERASEQIYYENRQGYRTASVITATDGHEDHDLFFYSEREANRSDLGAIYVCVG 180
Db 121 HEGFERASEQIYYENRQGYRTASVITATDGHEDHDLFFYSEREANRSDLGAIYVCVG 178
Qy 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHSLILKKSCEIILAAEPSTICAGESFQV 240
Db 179 KDFNETQLARIADSKDHVPVNDGFOALQGIHSLILKKSCEIILAAEPSTICAGESFQV 238
Qy 241 VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 300
Db 239 VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSDTYLLCPAPILKEVGMKAALQVSMN 298
Qy 301 DGLSFSSSVIITTHCS 321
Db 299 DGLSFSSSVIITTHCS 319

RESULT 13
```

US-10-038-307-16  
; Sequence 16, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 543  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-038-307-16

Query Match 85.1%; Score 1629; DB 6; Length 543;  
Best Local Similarity 98.8%; Pred. No. 6e-142;  
Matches 317; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

Qy 1 MATAERRALGIGFOWLSLATLVLCAGOGGREDGGPACYGFDLYFILDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFOWLSLATLVLCAGOGGREDGGPACYGFDLYFILDKSGSVLHHWN 60  
Qy 61 EIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Db 61 EIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Qy 121 HEGFERASEQIYYENRQGYRTASVIITDGLHEDLFFYSEREANRSDLGAIYVCVGV 180  
Db 121 HEGFERASEQIYYENRQGYRTASVIITDGLHEDLFFYSEREANRSDLGAIYVCVGV 178  
Qy 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHHSILKSKCIEILAAEPSTICAGESFOV 240  
Db 179 KDFNETQLARIADSKDHVPVNDGFOALQGIHHSILKSKCIEILAAEPSTICAGESFOV 238  
Qy 241 VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSEDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 239 VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSEDTYLLCPAPILKEVGMKAALQVSMN 298  
Qy 301 DGLSFSSSVIIITTHCSGKS 321  
Db 299 DGLSFSSSVIIITTHSSPKS 319

RESULT 15  
US-10-038-307-12  
; Sequence 12, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 534  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-038-307-12

Query Match 77.7%; Score 1488; DB 6; Length 534;  
Best Local Similarity 93.9%; Pred. No. 6.3e-129;  
Matches 294; Conservative 3; Mismatches 10; Indels 6; Gaps 2;

Qy 9 LGIGFOWLSLATLVLCAGOGGREDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQ 68  
Db 4 LGLLFCLVTLPCNL----SGRRREDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQ 59  
Qy 69 LAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYMHGFERAS 128  
Db 60 LAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYMHGFERAS 119  
Qy 129 EQIYYENRQGYRTASVIITDGLHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQL 188  
Db 120 EQIYYENRQGYRTASVIITDGLHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQL 177  
Qy 189 ARIADSKDHVPVNDGFOALQGIHHSILKSKCIEILAAEPSTICAGESFOVVVRNGFRH 248  
Db 178 ARIADSKDHVPVNDGFOALQGIHHSILKSKCIEILAAEPSTICAGESFOVVVRNGFRH 237  
Qy 249 ARNVDRLVLCSEKINDSVTLNEKPFVSEDTYLLCPAPILKEVGMKAALQVSMNDGLSFSS 308  
Db 238 ARNVDRLVLCSEKINDSVTLNEKPFVSEDTYLLCPAPILKEVGMKAALQVSMNDGLSFSS 297  
Qy 309 SVLIITTHCSGKS 321  
Db 299 SVLIITTHCSGKS 319

US-10-038-307-16  
; Sequence 16, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 543  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-038-307-16

Query Match 85.1%; Score 1629; DB 6; Length 543;  
Best Local Similarity 98.8%; Pred. No. 6e-142;  
Matches 317; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

Qy 1 MATAERRALGIGFOWLSLATLVLCAGOGGREDGGPACYGFDLYFILDKSGSVLHHWN 60  
Db 1 MATAERRALGIGFOWLSLATLVLCAGOGGREDGGPACYGFDLYFILDKSGSVLHHWN 60  
Qy 61 EIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Db 61 EIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDREQIROGLEELQKVLPGGDTYM 120  
Qy 121 HEGFERASEQIYYENRQGYRTASVIITDGLHEDLFFYSEREANRSDLGAIYVCVGV 180  
Db 121 HEGFERASEQIYYENRQGYRTASVIITDGLHEDLFFYSEREANRSDLGAIYVCVGV 178  
Qy 181 KDFNETQLARIADSKDHVPVNDGFOALQGIHHSILKSKCIEILAAEPSTICAGESFOV 240  
Db 179 KDFNETQLARIADSKDHVPVNDGFOALQGIHHSILKSKCIEILAAEPSTICAGESFOV 238  
Qy 241 VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSEDTYLLCPAPILKEVGMKAALQVSMN 300  
Db 239 VRNGFRHARNVDRVLCSEKINDSVTLNEKPFVSEDTYLLCPAPILKEVGMKAALQVSMN 298  
Qy 301 DGLSFSSSVIIITTHCSGKS 321  
Db 299 DGLSFSSSVIIITTHSSPKS 319

RESULT 14  
US-10-038-307-10  
; Sequence 10, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 543  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-038-307-10

Query Match 84.6%; Score 1619; DB 6; Length 543;  
Best Local Similarity 98.4%; Pred. No. 5.1e-141;  
Matches 316; Conservative 0; Mismatches 3; Indels 2; Gaps 1;



Db 298 SVIITTHSSPKS 310

Search completed: August 9, 2002, 10:35:55  
Job time: 353 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:30:01 ; Search time 45.48 seconds  
(without alignments)  
197.639 Million cell updates/sec

Title: US-09-970-076-2  
Perfect score: 1914  
Sequence: 1 MATARRALGIGFQWLSLAT.....VVIKEVPPPPABESENKIK 368

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Issued\_Patents\_AA.\*
- 1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*
  - 2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*
  - 3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*
  - 4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep.\*
  - 6: /cgn2\_6/ptodata/2/1aa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	166	8.7	1155	1 US-08-286-889-46	Sequence 46, Appl
2	166	8.7	1155	1 US-08-485-618-46	Sequence 46, Appl
3	166	8.7	1155	1 US-08-362-652-46	Sequence 46, Appl
4	166	8.7	1155	2 US-08-605-672-46	Sequence 46, Appl
5	166	8.7	1155	2 US-08-482-293A-46	Sequence 46, Appl
6	166	8.7	1155	2 US-08-943-363-46	Sequence 46, Appl
7	166	8.7	1155	4 US-09-193-043-46	Sequence 46, Appl
8	166	8.7	1161	1 US-08-485-618-53	Sequence 53, Appl
9	166	8.7	1161	2 US-08-362-652-53	Sequence 53, Appl
10	166	8.7	1161	2 US-08-605-672-53	Sequence 53, Appl
11	166	8.7	1161	2 US-08-482-293A-53	Sequence 53, Appl
12	166	8.7	1161	2 US-08-943-363-53	Sequence 53, Appl
13	166	8.7	1161	4 US-09-193-043-53	Sequence 53, Appl
14	159.5	8.3	1151	1 US-08-286-889-37	Sequence 37, Appl
15	159.5	8.3	1151	1 US-08-485-618-37	Sequence 37, Appl
16	159.5	8.3	1151	1 US-08-362-652-37	Sequence 37, Appl
17	159.5	8.3	1151	1 US-08-605-672-37	Sequence 37, Appl
18	159.5	8.3	1151	2 US-08-482-293A-37	Sequence 37, Appl
19	159.5	8.3	1151	2 US-08-943-363-37	Sequence 37, Appl
20	159.5	8.3	1151	4 US-09-193-043-37	Sequence 37, Appl
21	159.5	8.3	1161	1 US-08-173-497-2	Sequence 2, Appl
22	159.5	8.3	1161	1 US-08-286-889-2	Sequence 2, Appl
23	159.5	8.3	1161	1 US-08-485-618-2	Sequence 2, Appl
24	159.5	8.3	1161	1 US-08-485-618-55	Sequence 55, Appl
25	159.5	8.3	1161	1 US-08-485-618-99	Sequence 99, Appl
26	159.5	8.3	1161	1 US-08-362-652-2	Sequence 2, Appl
27	159.5	8.3	1161	1 US-08-362-652-55	Sequence 55, Appl

28	159.5	8.3	1161	2 US-08-605-672-2	Sequence 2, Appl
29	159.5	8.3	1161	2 US-08-605-672-55	Sequence 55, Appl
30	159.5	8.3	1161	2 US-08-605-672-99	Sequence 99, Appl
31	159.5	8.3	1161	2 US-08-482-293A-2	Sequence 2, Appl
32	159.5	8.3	1161	2 US-08-482-293A-55	Sequence 55, Appl
33	159.5	8.3	1161	2 US-08-482-293A-99	Sequence 99, Appl
34	159.5	8.3	1161	2 US-08-943-363-2	Sequence 2, Appl
35	159.5	8.3	1161	2 US-08-943-363-55	Sequence 55, Appl
36	159.5	8.3	1161	2 US-08-943-363-99	Sequence 99, Appl
37	159.5	8.3	1161	4 US-09-193-043-2	Sequence 2, Appl
38	159.5	8.3	1161	4 US-09-193-043-55	Sequence 55, Appl
39	159.5	8.3	1161	4 US-09-193-043-99	Sequence 99, Appl
40	149.5	7.8	435	5 PCT-US95-04439-1	Sequence 1, Appl
41	145.5	7.6	1152	2 US-08-476-062A-43	Sequence 43, Appl
42	145.5	7.6	1152	5 PCT-US96-01314-43	Sequence 43, Appl
43	145.5	7.6	1152	6 5424399-2	Patent No. 5424399
44	145.5	7.6	1153	1 US-08-173-497-3	Sequence 3, Appl
45	145.5	7.6	1153	1 US-08-286-889-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1  
US-08-286-889-46  
; Sequence 46, Application US/08286889  
; Patent No. 5470953  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Mich  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: NO. 5470953el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/286.889  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: P38,659  
; REFERENCE/POCKET NUMBER: 27866/32168  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-286-889-46

Query Match 8.7%; Score 166; DB 1; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 4.2e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;  
QY 37 PACYGFGLYFKSGSV-LHHWNIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMK 94

Db 144 PECPGQEMDIAFLIDSGSIDQSDFTQMKDFVKALMGOLASTSTSFSLMQYSNLIKTHFT 203  
QY 95 LTEDR-----EQIROGLEELQKVLPGDGYMHGFEFRASEQIYYENRQGYRTA-SVIAL 148  
Db 204 FTEFKSSLSQSLVDAIVQLQ-----GLTYTASGIQKVVKELFHSKNGARKSAKILIVI 258  
QY 149 TDGELHEDLFFYSE--REANRSRDGAIVYCVGKVD-FNE-TOLARI-----ADSKDHVF 199  
Db 259 TDGQKFRDPLEYRHVPEAKA---GIIRYAIGVGDAFREPTALQELNLTIGSAPSDHVF 315  
QY 200 PVNDGFOALOGIHSILKKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVD 253  
Db 316 KVGN-FVALRSIQRIQOEK-----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

RESULT 3  
US-08-362-652-46  
; Sequence 46, Application US/08362652  
; Patent No. 5766850  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/362,652  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32391  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-362-652-46

Query Match 8.7%; Score 166; DB 1; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 4.2e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 37 PACYG-GFDLYFLDKSGSV-LHHWNEIYFVEQLAHKFIQPLRMFSFIVSTRGTTLMK 94  
Db 144 PECPGQEMDIAFLIDSGSIDQSDFTQMKDFVKALMGOLASTSTSFSLMQYSNLIKTHFT 203  
QY 95 LTEDR-----EQIROGLEELQKVLPGDGYMHGFEFRASEQIYYENRQGYRTA-SVIAL 148

Db 144 PECPGQEMDIAFLIDSGSIDQSDFTQMKDFVKALMGOLASTSTSFSLMQYSNLIKTHFT 203  
QY 95 LTEDR-----EQIROGLEELQKVLPGDGYMHGFEFRASEQIYYENRQGYRTA-SVIAL 148  
Db 204 FTEFKSSLSQSLVDAIVQLQ-----GLTYTASGIQKVVKELFHSKNGARKSAKILIVI 258  
QY 149 TDGELHEDLFFYSE--REANRSRDGAIVYCVGKVD-FNE-TOLARI-----ADSKDHVF 199  
Db 259 TDGQKFRDPLEYRHVPEAKA---GIIRYAIGVGDAFREPTALQELNLTIGSAPSDHVF 315  
QY 200 PVNDGFOALOGIHSILKKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVD 253  
Db 316 KVGN-FVALRSIQRIQOEK-----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

RESULT 2  
US-08-485-618-46  
; Sequence 46, Application US/08485618  
; Patent No. 5728533  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485,618  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32797  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-485-618-46

Query Match 8.7%; Score 166; DB 1; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 4.2e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 37 PACYG-GFDLYFLDKSGSV-LHHWNEIYFVEQLAHKFIQPLRMFSFIVSTRGTTLMK 94

		Query Match	8.7%	Score 166;	DB 2;	Length 1155;
		Best Local Similarity	28.2%;	Pred. No. 4.2e-09;		
		Matches	66;	Conservative 42;	Mismatches 96;	Indels 30; Gaps
Qy	37	PACVG-GFDLYFLDKSGSV-LHHWNEIYFVQLAHKFTSPOLRMSFIVFSTRGTTLMK	94			
		:   :   :   :   :   :   :   :   :   :   :   :   :   :				
Db	144	PECPQEMDIAFLIDGSGSIDQSDFTQMOKDFVALMQLASTSTSFSLMGSNILKTHFT	203			
Qy	95	LTERD-----EQIRQLEELQVLPGGDTYMHGFEFRASEQIYYENRGVRPTA-SVIAL	148			
	:	:   :   :   :   :   :   :   :   :   :   :   :   :   :				
Db	204	FTEFKSLSPOSVDATVQLQ-----GLTDTAGIQRVVKLFPSHKNARKSAKKILIVI	258			

	Query Match	8.7%	Score 166;	DB 2;	Length 1155;
	Best Local Similarity	28.2%	Pred. No. 4.2e-09;		
	Matches 66;	Conservative 42;	Mismatches 96;	Indels 30;	Gaps
Qy	37	PACYG-GEDLYFILDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFIYFVSTRGITLMK	94		
Db	144	PECPQEMDIAFLIDGSGSIDQSDFTQMKDFVKALMGQLASTSTSFSLMOYSNILKTHET	203		
Qy	95	LTEDR-----EQIRQGLEELQKVLPGGDTYMHGCFRASEQIYVENRQGYRTA-SVIAAL	148		
Db	204	FTFRKSSLPQSLDAIVQLQ-----GLTTFASGIQKRWKLELHFKSNKGARSKAKKILIVI	258		

```

Query Match      8.7%; Score 166; DB 2; Length 1155;
Best Local Similarity 28.2%; Pred. No. 4.2e-09;
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

37 PACYGFGLYFILDKGSV-LHHWNEIYFVEQLAHKFIISPOLRMSFVIFSTRGTTLMK 94
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
144 PCPCGEMDIALIDGSGISDOSDTQKDFKALMGQLASTSTSFSLMQYSNLIKTHFT 203

95 LTEDR-----EOTRGGLEELQVLPGGDTYMHGFERASEQIYYENRGYRTA-SVIAL 148
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
204 FTEFKSSLSPQSLDAIVQLQ-----GLTYTASGQKVYKELFHSHKNGARKSAKKILIVI 258

```



; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485,618  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32797  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1161 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-485-618-53

Query Match 8.7%; Score 166; DB 1; Length 1161;  
Best Local Similarity 28.2%; Pred. No. 4.3e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 37 PACYG-GFDLYFILDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFVFTSGTTLMK 94  
DB 144 PCPCQGMEDIAFLDGSIDQSDFTQMKDFVKALMGQLASTSTSFSLMOYSNLIKTHFT 203  
QY 95 LTEDR-----EOIRGLELQKVLPGDYMHEGFERASEQIYYENRQYRTA-SVILAL 148  
DB 204 FTEFKSSLSQSLVDAIVQLQ-----GLTYTAGIQKVVKELFHSKNGARKSAKKILIVI 258  
QY 149 TDGELHEDLFFYSE--REANRSRLGAIVYCVGVKD-FNE-TOLARI-----ADSKDHVF 199  
DB 259 TDGQKFRDPLRYRHVPEAKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSODHVF 315  
QY 200 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFOVVYRNGFRHARNVD 253  
DB 316 KVG-N-FVALRSIQRIQEK-----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

RESULT 9  
US-08-652-53  
; Sequence 53, Application US/08362652  
; Patent No. 5766850  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/362,652  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32391  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1161 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-362-652-53

Query Match 8.7%; Score 166; DB 1; Length 1161;  
Best Local Similarity 28.2%; Pred. No. 4.3e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 37 PACYG-GFDLYFILDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFVFTSGTTLMK 94  
DB 144 PCPCQGMEDIAFLDGSIDQSDFTQMKDFVKALMGQLASTSTSFSLMOYSNLIKTHFT 203  
QY 95 LTEDR-----EOIRGLELQKVLPGDYMHEGFERASEQIYYENRQYRTA-SVILAL 148  
DB 204 FTEFKSSLSQSLVDAIVQLQ-----GLTYTAGIQKVVKELFHSKNGARKSAKKILIVI 258  
QY 149 TDGELHEDLFFYSE--REANRSRLGAIVYCVGVKD-FNE-TOLARI-----ADSKDHVF 199  
DB 259 TDGQKFRDPLRYRHVPEAKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSODHVF 315  
QY 200 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFOVVYRNGFRHARNVD 253  
DB 316 KVG-N-FVALRSIQRIQEK-----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

RESULT 10  
US-08-605-672-53  
; Sequence 53, Application US/08605672  
; Patent No. 5817515  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

us-09-970-076-2.ra1

Fri Aug 9 10:56:59 2002

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,293A
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-482-293A-53

```

```

Query Match      8.7%; Score 166; DB 2; Length 1161;
Best Local Similarity 28.2%; Pred. No. 4.3e-09;
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

```

```

;
; Query Match      8.7%; Score 166; DB 2; Length 1161;
; Best Local Similarity 28.2%; Pred. No. 4.3e-09;
; Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;
;
; QY 37 PACYG-GFDLYFLDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFVIFSTRGTTLMK 94
; DB 144 PCCPGQEMDIAFLIDGSGSIDQSDFTQMKDFVKALMGQLASTSTFSLMQYSNILKTHFT 203
; QY 95 LTEDR-----EQIRQGLEELQKVLPGDGYMHGEPERASEQIYYENRQGYRTA-SVIAL 148
; DB 204 FTEFKSSLSQSLVDAIVOLQ-----GLIYTASGIQKVVKELFHSKNGARKSAKKILIVI 258
; QY 149 TDGELHEDLFFYSE--REANRSRDLGAIVYCVGVKD-FNE-TQLARI-----ADSKDHVF 199
; DB 259 TDGQKFRDPLEYRHVPEAKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSQDHVF 315
; QY 200 PVNDGQALQGIHLSILKSCIEILAAEPSTICAGESFQVVRGNGFRHARNVD 253
; DB 316 KVG N-FVALRSIQRIQEK----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

```

```

RESULT 11
US-08-482-293A-53
; Sequence 53, Application US/08482293A
; Patent No. 5831029
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/605,672
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-605-672-53

```

```

Query Match      8.7%; Score 166; DB 2; Length 1161;
Best Local Similarity 28.2%; Pred. No. 4.3e-09;
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

```

```

;
; QY 37 PACYG-GFDLYFLDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFVIFSTRGTTLMK 94
; DB 144 PCCPGQEMDIAFLIDGSGSIDQSDFTQMKDFVKALMGQLASTSTFSLMQYSNILKTHFT 203
; QY 95 LTEDR-----EQIRQGLEELQKVLPGDGYMHGEPERASEQIYYENRQGYRTA-SVIAL 148
; DB 204 FTEFKSSLSQSLVDAIVOLQ-----GLIYTASGIQKVVKELFHSKNGARKSAKKILIVI 258
; QY 149 TDGELHEDLFFYSE--REANRSRDLGAIVYCVGVKD-FNE-TQLARI-----ADSKDHVF 199
; DB 259 TDGQKFRDPLEYRHVPEAKA---GIIRYAIGVGDAFREPTALQELNTIGSAPSQDHVF 315
; QY 200 PVNDGQALQGIHLSILKSCIEILAAEPSTICAGESFQVVRGNGFRHARNVD 253
; DB 316 KVG N-FVALRSIQRIQEK----IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

```

```

RESULT 11
US-08-482-293A-53
; Sequence 53, Application US/08482293A
; Patent No. 5831029
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```



; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32684  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1161 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-943-363-53

Query Match 8.7%; Score 166; DB 2; Length 1161;  
Best Local Similarity 28.2%; Pred. No. 4.3e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 37 PACYGFGLDYLKSGSV-LHHWNIYYFVEQLAHKFISPOLRMSFIVSTRGTTLMK 94  
DB 144 PCPCQEMDIAFLIDGSGIDQSDFTQMKDFVKALMGQLASTSFSLMQVSNILKTHFT 203  
QY 95 LTEDR-----EQIROGLELQKVLPGGDTYMHGEPERASEQIYYENRGYRTA-SVIAL 148  
DB 204 FTEFKSSLSQSLVDIAVLQ-----GLITYASGLQKVVKELFHSKNGARKSAKKILIVI 258  
QY 149 TDGELHEDLFFYSE--REANKSRDLGAIVYCVGVKD-FNE-TQLARI-----ADSKDHVF 199  
DB 259 TDGQKFRDPLEYRHRVPIPEAKA--GIIRYAIGVGDFAFREPALQELNTIGSAPSQDHVF 315  
QY 200 PVNDGFQALQGLIIHSLKSCIEILAAEPSTICAGESFQVVVGRNGFRHARNVD 253  
DB 316 KVGNFVALRSIQRIQEK----IFAIEGTERSSSFQHEMSQEGFSSALSMD 364

RESULT 13  
US-09-193-043-53  
; Sequence 53, Application US/09193043  
; Patent No. 6251395  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, Michael W.  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 6251395el Human 2  
; FILE REFERENCE: 27866/35004  
; CURRENT APPLICATION NUMBER: US/09/193,043  
; CURRENT FILING DATE: 1998-11-16  
; EARLIER APPLICATION NUMBER: 08/173,497  
; EARLIER FILING DATE: 1993-12-23  
; EARLIER APPLICATION NUMBER: 08/286,889  
; EARLIER FILING DATE: 1994-08-05  
; EARLIER APPLICATION NUMBER: 08/362,652  
; EARLIER FILING DATE: 1994-12-21  
; EARLIER APPLICATION NUMBER: 08/943,363  
; EARLIER FILING DATE: 1997-10-03  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 53  
; LENGTH: 1161  
; TYPE: PRT

; ORGANISM: Mus musculus  
US-09-193-043-53

Query Match 8.7%; Score 166; DB 4; Length 1161;  
Best Local Similarity 28.2%; Pred. No. 4.3e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 37 PACYGFGLDYLKSGSV-LHHWNIYYFVEQLAHKFISPOLRMSFIVSTRGTTLMK 94  
DB 144 PCPCQEMDIAFLIDGSGIDQSDFTQMKDFVKALMGQLASTSFSLMQVSNILKTHFT 203  
QY 95 LTEDR-----EQIROGLELQKVLPGGDTYMHGEPERASEQIYYENRGYRTA-SVIAL 148  
DB 204 FTEFKSSLSQSLVDIAVLQ-----GLITYASGLQKVVKELFHSKNGARKSAKKILIVI 258  
QY 149 TDGELHEDLFFYSE--REANKSRDLGAIVYCVGVKD-FNE-TQLARI-----ADSKDHVF 199  
DB 259 TDGQKFRDPLEYRHRVPIPEAKA--GIIRYAIGVGDFAFREPALQELNTIGSAPSQDHVF 315  
QY 200 PVNDGFQALQGLIIHSLKSCIEILAAEPSTICAGESFQVVVGRNGFRHARNVD 253  
DB 316 KVGNFVALRSIQRIQEK----IFAIEGTERSSSFQHEMSQEGFSSALSMD 364

RESULT 14  
US-08-286-889-37  
; Sequence 37, Application US/08286889  
; Patent No. 5470953  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Mich  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/286,889  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32168  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 37:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1151 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-286-889-37

Query Match 8.3%; Score 159.5; DB 1; Length 1151;  
Best Local Similarity 27.9%; Pred. No. 2.2e-08;  
Matches 63; Conservative 41; Mismatches 93; Indels 29; Gaps 11;



---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:32:17 ; Search time 102.68 seconds  
(without alignments)  
620.005 Million cell updates/sec

Title: US-09-970-076-2  
Perfect score: 1914  
Sequence: 1 MATAERRALGIGFQWLSLAT.....VLIKEVPPPPAESEENKIK 368

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 17294929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: SP TREMBL\_19:\*
- 2: sp.archaea:\*
- 3: sp.bacteria:\*
- 4: sp.fungi:\*
- 5: sp.human:\*
- 6: sp.invertebrate:\*
- 7: sp.mammal:\*
- 8: sp.mhc:\*
- 9: sp.organelle:\*
- 10: sp.phage:\*
- 11: sp.plant:\*
- 12: sp.rodent:\*
- 13: sp.virus:\*
- 14: sp.vertibrate:\*
- 15: sp.unclassified:\*
- 16: sp.rvirus:\*
- 17: sp.bacteriap:\*
- 17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1914	100.0	368	4 Q96P02	Q96P02 homo sapien
2	574	30.0	245	4 Q96NC7	Q96NC7 homo sapien
3	391.5	20.5	97	4 Q96EC6	Q96EC6 homo sapien
4	286.5	15.0	221	11 Q99L17	Q99L17 mus musculus
5	159.5	8.3	1161	11 Q90VE7	Q90VE7 rattus norv
6	153.5	8.0	1332	5 Q9BP08	Q9BP08 halocynthia
7	150.5	7.9	724	5 Q04588	Q04588 elmeria max
8	137.5	7.2	712	5 Q43981	Q43981 elmeria ten
9	134	7.0	760	11 Q70350	Q70350 mus musculus
10	132.5	6.9	765	5 Q9U8J9	Q9U8J9 neospora ca
11	131.5	6.9	517	4 Q43853	Q43853 homo sapien
12	131	6.8	660	4 Q9UC3	Q9UC3 homo sapien
13	130.5	6.8	920	5 Q28984	Q28984 sus scrofa
14	128	6.7	3767	5 Q9UAI3	Q9UAI3 caenorhabdi
15	123	6.4	1151	11 Q9J130	Q9J130 rattus norv
16	118.5	6.2	3567	11 Q9ES77	Q9ES77 mus musculus

17	117.5	6.1	1140	4 Q9P218	Q9P218 homo sapien
18	117.5	6.1	1207	4 Q9BOU7	Q9BOU7 homo sapien
19	117.5	6.1	2944	11 Q63870	Q63870 mus musculus
20	114.5	6.0	2104	5 Q21281	Q21281 caenorhabdi
21	114.5	6.0	2104	5 Q964N4	Q964N4 caenorhabdi
22	113.5	5.9	755	4 Q00261	Q00261 homo sapien
23	113.5	5.9	764	4 Q96HX6	Q96HX6 homo sapien
24	113.5	5.9	937	4 Q96FT5	Q96FT5 homo sapien
25	113	5.9	833	13 Q9VIC5	Q9VIC5 cyprinus ca
26	112	5.9	650	11 Q9CY21	Q9CY21 mus musculus
27	111	5.8	1905	5 Q9XTP6	Q9XTP6 plasmodium
28	109.5	5.7	704	10 Q9FF49	Q9FF49 arabidopsis
29	109	5.7	1253	6 Q97566	Q97566 canis famil
30	108.5	5.7	956	11 Q99K64	Q99K64 mus musculus
31	107	5.6	747	13 Q91900	Q91900 xenopus lae
32	106.5	5.6	562	5 Q01510	Q01510 plasmodium
33	106	5.5	415	11 Q923K3	Q923K3 mus musculus
34	106	5.5	833	5 Q96442	Q96442 strongyloce
35	105.5	5.5	567	5 Q18048	Q18048 caenorhabdi
36	104.5	5.5	769	5 Q00816	Q00816 toxoplasma
37	104.5	5.5	1169	11 Q9QXH4	Q9QXH4 mus musculus
38	104	5.4	205	11 Q63001	Q63001 rattus norv
39	102.5	5.4	537	4 Q96AA0	Q96AA0 homo sapien
40	102	5.3	614	5 Q94674	Q94674 plasmodium
41	102	5.3	714	10 Q948G7	Q948G7 oryza sativ
42	102	5.3	759	4 Q96C67	Q96C67 homo sapien
43	101.5	5.3	425	5 Q9GZF5	Q9GZF5 caenorhabdi
44	101.5	5.3	559	5 Q01506	Q01506 plasmodium
45	101.5	5.3	559	5 Q01508	Q01508 plasmodium

## ALIGNMENTS

RESULT 1

Q96P02	ID	Q96P02	PRELIMINARY;	PRT;	368 AA.
AC	Q96P02;				
DT	01-DEC-2001	(Tremblrel. 19, Created)			
DT	01-DEC-2001	(Tremblrel. 19, Last sequence update)			
DT	01-DEC-2001	(Tremblrel. 19, Last annotation update)			
DE	ANTHRAX TOXIN RECEPTOR.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Bradley K.A., Moggridge J., Mourez J., Collier R.J., Young J.A.T.;				
RT	"Identification of the Cellular Receptor for Anthrax Toxin.";				
RL	Nature 414:0-0(2001).				
DR	EMBL; AF421380; AAL26496.1;				
KW	Receptor.				
SK	SEQUENCE 368 AA; 41157 MW; 8A87B13FFA7D8753 CRC64;				

Query Match	100.0%	Score 1914;	DB 4;	Length 368;
Best Local Similarity	100.0%;	Pred. No. 7.6e-160;		
Matches 368;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MATAERRALGIGFQWLSLATLVLCAGQGRREDGGPACYGFDLYFLDKSGSVLHHWN	60	
Db	1	MATAERRALGIGFQWLSLATLVLCAGQGRREDGGPACYGFDLYFLDKSGSVLHHWN	60	
Qy	61	EIYFVQLAHKFTSPQLRMSFIVFSTRGTTLMKLTEDREQIRGLELQVLPGGDTYM	120	
Db	61	EIYFVQLAHKFTSPQLRMSFIVFSTRGTTLMKLTEDREQIRGLELQVLPGGDTYM	120	
Qy	121	HEGPERASEQIYYENRGYRTASVITATDGHEDLFFYSERANRSDLGATVYCVGV	180	
Db	121	HEGPERASEQIYYENRGYRTASVITATDGHEDLFFYSERANRSDLGATVYCVGV	180	
Qy	181	KDFNETQLARIADSKDHFVPVNDGFGALQGIHSLTKKSCIEILAAEPSTICAGESFOV	240	

Fri Aug 9 10:57:01 2002

Db 181 KDFNETQIARIADSKDHVPVNDGFOALOGIIHSILKKSCIEILAAEPSTICAGESFQV 240  
 QY 241 VRNGFRHARNVDRVLCSPKINDSVTLNEKPFVSVDYLLCPAPILKEVGMKAALQVSN 300  
 Db 241 VRNGFRHARNVDRVLCSPKINDSVTLNEKPFVSVDYLLCPAPILKEVGMKAALQVSN 300  
 QY 301 DGLSFTSSSVIITTHCSGSLAIALLILFLLALLLWFWPLCCTVIKEVPPPAE 360  
 Db 301 DGLSFTSSSVIITTHCSGSLAIALLILFLLALLLWFWPLCCTVIKEVPPPAE 360  
 QY 361 ESEBENKIK 368  
 Db 361 ESEBENKIK 368

RESULT 2  
 Q96NC7 PRELIMINARY; PRT; 245 AA.  
 AC Q96NC7;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE CDNA FLJ31074 FIS, CLONE HSYRA2001476.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 PP SEQUENCE FROM N.A.  
 RA Ninomiya K., Wagatsuma M., Kikawa E., Kondo H., Yokoi T., Kodaira H.,  
 RA Furuya T., Takahashi M., Kikawa E., Omura Y., Abe K., Kamihara K.,  
 RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,  
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,  
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
 RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;  
 RA "NEDO human cDNA sequencing project";  
 RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK05636; BAB70976.1; -;  
 SQ SEQUENCE 245 AA; 26111 MW; 61AE6EB0A2EBEE06 CRC64;

Query Match 30.0%; Score 574; DB 4; Length 245;  
 Best Local Similarity 49.2%; Pred. No. 2.2e-42;  
 Matches 120; Conservative 45; Mismatches 67; Indels 12; Gaps 3;  
 QY 78 LRMSFTVSTRTTLMKLTEDREQIQLGELQVLPDGTVMHGFSEFIYENRQ 137  
 Db 1 MRLSFTVSSQATILPLTGDGRKISKGLDLKRVSPVGETYIHEGLKLANEQI--QKAG 58  
 QY 138 GYRTASVIALTDGELHEDLFYSEREANRSDLGAIYVCVGVKDFNETQLARIADSKDH 197  
 Db 59 GLKTSIIIALTDGKLDGULGVPSYAKEAKISRLGASVCVGVLDPEQAQLERIADSKQE 118  
 QY 198 VFPVNDGFOALOGIIHSILKSCIEILAAEPSTICAGESFQVVRNGFRHARNVDRVLC 257  
 Db 119 VFPVKGFOALGIINSILAQSCITEILELPQSSVCVGVGEFVLGSRGFMLSRNGSLVC 178  
 QY 258 SFKINDSVTLNEKPFVSVDYLLCPAPILKEVGMKAALQVSNMNDGLSFSTSSV-IITTH 316  
 Db 179 TYTVNETYTSVKPVSQVLSNMLCPAPILNKAG-----EWGLTVTQAGVKHDLTH 229  
 QY 317 CSOG 320  
 Db 230 CTFG 233

RESULT 3  
 Q96EC6 PRELIMINARY; PRT; 97 AA.  
 AC Q96EC6;

DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE UNKNOWN (PROTEIN FOR IMAGE:4705862) (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 PP SEQUENCE FROM N.A.  
 RC TISSUE-BREAST, AND MAMMARY ADENOCARCINOMA;  
 RA Strausberg R.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC012475; AAH12475.1; -;  
 FT NON\_TER 1  
 SQ SEQUENCE 97 AA; 10453 MW; 14F475F0B170E71A CRC64;

Query Match 20.5%; Score 391.5; DB 4; Length 97;  
 Best Local Similarity 90.0%; Pred. No. 6.6e-27;  
 Matches 81; Conservative 1; Mismatches 5; Indels 3; Gaps 1;  
 QY 237 FOVVVRNGFRHARNVDRVLCSPKINDSVTLNEKPFVSVDYLLCPAPILKEVGMKAALQ 296  
 Db 1 FOVVVRNGFRHARNVDRVLCSPKINDSVTLNEKPFVSVDYLLCPAPILKEVGMKAALQ 60  
 QY 297 VSMNDGLSFISSVIITTHCSGSLAIATA 326  
 Db 61 VSMNDGLSFISSVIITTHCSGSLAIATA 87

RESULT 4  
 Q99L17 PRELIMINARY; PRT; 221 AA.  
 ID Q99L17;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE HYPOTHETICAL 24.5 KDA PROTEIN (FRAGMENT).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 PP SEQUENCE FROM N.A.  
 RC TISSUE-MAMMARY TUMOR. WAP-TGF ALPHA MODEL. 7 MONTHS OLD, GROSS  
 RC TISSUE.;  
 RA Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC003908; AAH03908.1; -;  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 221 AA; 24505 MW; 31AB702E42E9CEB3 CRC64;

Query Match 15.0%; Score 286.5; DB 11; Length 221;  
 Best Local Similarity 49.0%; Pred. No. 3.4e-17;  
 Matches 50; Conservative 21; Mismatches 28; Indels 3; Gaps 1;  
 QY 269 EKPSVVDYLLCPAPILKEVGMKAALQVSNMNDGLSFISSVIITTHCSGSLAIATA 328  
 Db 1 EKPSVQDSSLLCPAPLNKDGTELVSIYNDGKSAVSRSLTITATECTINGIAAIVAIL 60  
 QY 329 ILFLLALLLWFWPLCCTVIKEVPPPA---EESSEENKI 367  
 Db 61 VLLLLGALLWFWPLCKVKVVKDPDPPPPSAPMEEEEDPL 102

RESULT 5  
 Q9QYE7 PRELIMINARY; PRT; 1161 AA.  
 ID Q9QYE7  
 AC Q9QYE7;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)

DR EMBL; AB048261; BAB21479.1; -.  
DR HSSP; P11215; IA8X.  
DR InterPro; IPR000413; Integrin\_alpha.  
DR InterPro; IPR002035; VWFA.  
DR Pfam; PF01839; FG-GAP; 5.  
DR Pfam; PF00357; integrin\_A; 1.  
DR Pfam; PF00092; vwa; 1.  
DR PRINTS; PR01185; INTEGRINA.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00191; Int\_alpha; 5.  
DR SMART; SM00327; VWA; 1.  
DR PROSITE; PS50234; VWFA; 1.  
KW Signal.  
FT SIGNAL  
FT CHAIN 1 30 POTENTIAL.  
FT CHAIN 31 1332 INTEGRIN ALPHA HR1.  
SQ SEQUENCE 1332 AA; 145852 MW; 0D9108D2B05CFFAE CRC64;

Query Match 8.0%; Score 153.5; DB 5; Length 1332;  
Best Local Similarity 24.1%; Pred. No. 0.00019;  
Matches 63; Conservative 42; Mismatches 81; Indels 75; Gaps

QY	29	GGREDGG-----PACVCGGFDLYFLDKSGSVLHHNNEIYYFYFEQLAHKEISPOLRMSFI 83
DB	188	GNNRKGEGRNRTECPSS--SGVDVLFLVDGSGVGKFNDFKDVMKNIT----- 232
QY	84	VFSTRGTTLMKLTEDRSQIROGLEELQKVLPG-----GDTYMHGFERAS 138
DB	233	-----AKLDIGKEIVRGVVQSYHYEGKSINKQKYITTEISIGEKFLLDNENAV 263
QY	129	EQIYYENROGYRT-----ASVIAITDGELHEDLFYSEREANRS 168
DB	284	DRI---QLQGTYTTRALQKVIREFDDFVAYIGNKQVLLLLLTDOAKDNKLILP--NANRL 338
QY	169	RDLCAIYVCYGVKDFNETQLARIA---DSKHVPYPNDGFQAOLGIHSILKSKSCIEILA 225
DB	339	RNGKIATFAYGVGGDYISELKLIASGTSDTRVFTVTVD-FGELODSIVKSLQTEIQSFVLE 397
QY	226	AEPSTICAGESFOVVVRGNCF 246
DB	398	GGSKAKTAG--YEMHFGENCF 416

RESULT 7

ID	Q04588	PRELIMINARY;	PRT;	724 AA.
AC	Q04588:			
DT	01-NOV-1996	(TrEMBLrel. 01, Created)		
DT	01-NOV-1996	(TrEMBLrel. 01, Last sequence update)		
DE	01-DEC-2001	(TrEMBLrel. 19, Last annotation update)		
DE	MAJOR ANTIGEN HOMOLOGOUS SEQUENCE (EMP100).			
OS	Elmeria maxima.			
OC	Eukaryota; Alveolata; Apicomplexa; Coccidia; Elmeriida; Elmeriidae;			
CC	Elmeriota.			
OX	NCBI_TaxID=5804;			
ON	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE-93149203; PubMed-8426611;			
EX	Paramontes L.E.; Hug D.; Huembelin M., Weber G.;			
RT	"Sequence of a major Elmeria maxima antigen homologous to the Elmeriota			
RT	tellena microneme protein Etp100.";			
RL	Mol. Biochem. Parasitol. 57:171-174(1993).			
DR	EMBL; M95058; AAA29076.1; .			
DR	InterPro; IPR000884; TSP1.			
DR	InterPro; IPR002035; VWFA.			
DR	Pfam; PF00090; tsp_1; 6.			
DR	Pfam; PF00092; vwa; 1.			
DR	PRINTS; PR00453; VWFADOMAIN.			
DR	SMART; SM00209; TSP1; 6.			
DR	SMART; SM00327; VWA; 1.			
DR	PROSITE; PSS0092; TSP1; 5.			
DR	PROSITE; PSS0234; VWFA; 1.			
SQ	SEQUENCE 724 AA; 75808 MW; AC2A0E7A346A7B9E CRC64;			

Query Match		7.9%;	Score 150.5;	DB 5;	Length 724;
Best Local Similarity		25.9%;	Pred. No. 0.00015;		
Matches 69;		Conservative 45;	Mismatches 109;	Indels 43;	Gaps 17;
QY	1	MATAERRALGIGFOWLSLATLVLCAG-----QGRREDGPACYGFGDLYFILKSGS	54		
DB	1	MALLPTQRLAP--W-ALSLVFLAAGLTFHSSHAASSEADQVCTRLDVLVWVDES	57		
QY	55	V-LHHWNEIYFVQLAHKF-ISP-QLRMSFIVFSTGTTLMLKLTEDREIQROGLELOK	111		
DB	58	IGTSNYKVRFSFISNFACTWPLSPDDVVRVGLVTFGTSAVTRWDLSRAONADLLAAAK	117		
QY	112	VLP--GGDTYMHEGFERASEQIYYENRQGYR--TASVIIALTGDGELHEDLFFYSERE---	164		
DB	118	KLPTAAGSTYTHLGLAKA-EELFSFKGGRDNAPKMLVMTDGA-----SSRSQT	168		
QY	165	---ANRSDLGAIYVCVGV-KDFNETOLARIA--DSKDHV-FP--VNDGFOALOGIITHSI	215		
DB	169	LSAAEKLNRGVIIIVLGVGTGVNSAECRSIAGCDTSDTVECPRYLOSNGVSSQINGI	228		
QY	216	LKKSCIEI-----LAAEPSTI--CAGE 235			
DB	229	IKACKDLAKDAVCSEYSEYGPCEGE 254			
RESULT 8					
ID	043981	PRELIMINARY;	PRT;	712	AA.
AC	043981;				
DT	01-JUN-1998	(Tremblrel. 06, Created)			
DT	01-JUN-1998	(Tremblrel. 06, Last sequence update)			
DT	01-DEC-2001	(Tremblrel. 19, Last annotation update)			
DE	MICRONEME PROTEIN ETMIC-1 PRECURSOR.				
GN	MIC-1.				
OS	Eimeria tenella.				
OC	Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;				
OC	Eimeria.				
OX	NCBI_TaxID=5802;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=HOUGHTON;				
RX	MEDLINE=99094493;				
RA	Tomley F.M., Tomley F.M.; PubMed=1775171;				
RA	Tomley F.M., Clarke L.E., Kawazoe U., Dijkema R., Kok J.J.;				
RT	"Sequence of the gene encoding an immunodominant microneme protein of				
RT	Eimeria tenella."				
RL	Mol. Biochem. Parasitol. 97:21-31(1998).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=HOUGHTON;				
RX	MEDLINE=99094493;				
RA	Kelleher M., Tomley F.M.;				
RT	"Transient expression of beta-galactosidase in differentiating				
RT	sporozoites of Eimeria tenella."				
RL	Mol. Biochem. Parasitol. 97:21-31(1998).				
DR	EMBL;	AF032905;	AAD03350.1;		
DR	HSSP;	P17301;	1A0X.		
DR	InterPro;	IPR000884;	TSPL.		
DR	InterPro;	IPR002035;	VFPA.		
DR	Pfam;	PF00090;	tsp-1;	6.	
DR	Pfam;	PF00092;	vwa;	1.	
DR	PRINTS;	PR00453;	VWFA.		
DR	SMART;	SM00209;	TSPL;	6.	
DR	PROSITE;	PS50092;	TSPL;	5.	
DR	PROSITE;	PS50234;	VFPA;	1.	
KW	Signal.				
FT	SIGNAL	1	36	POTENTIAL.	
FT	CHAIN	37	712	MICRONEME PROTEIN ETMIC-1.	
SQ	SEQUENCE	712	AA;	74777	MW; 15B8F3C190B70F73

Query Match		7.2%;	Score 137.5;	DB 5;	Length 712;
Best Local Similarity		24.7%;	Pred. No. 0.0021;		
Matches 67;		Conservative 43;	Mismatches 108;	Indels 53;	Gaps 14;
QY	1	MATAERRALGIGFOWLSLATLVLCAG-----QGRREDGG--PACYGGFDLYFILDK	51		
DB	1	MAPLPRRLA--PCRALSLVGLLAASFAFSLQPGATFSSGQDOVCTSLDVLVWVDE	57		
QY	52	SGSV-LHHWNEIYFVQLAHKF-ISPO-LRMSFIVFSTGTTLMLKLTEDREIQROGLEE	108		
DB	58	SGSIGTSNFKRVQFTIEDFVNSMPISPEDVRVGLITFAI-----RSKVRNLS	106		
QY	109	LQKVLPG-----GDTYMHEGFERASEQIYYENRQGYR--TASVIIALTGDGEL	153		
DB	107	PKATNPSLAISAARSLSYSTGVTYHYGLQDA-KLLYDTNAGARNNVKLVMTDGA-	164		
QY	154	HEDLFYSREANRSDLGAIYVCVGV-KDFNETOLARIAADSKDHVP-----VNDGFO	206		
DB	165	-SNLPSQTRSSAAALRDAGAIIVVLGVGVSNSSECRSAGCSTSNCPRYLOSNSNVQ	223		
QY	207	ALOGIITHSILKSCIEILAAEPSTI--CAGE 235			
DB	224	QVNGIILKACKDLAKDAVCSEYSEYGPVCGE 254			
RESULT 9					
ID	070350	PRELIMINARY;	PRT;	760	AA.
AC	070350;				
DT	01-AUG-1998	(Tremblrel. 07, Created)			
DT	01-AUG-1998	(Tremblrel. 11, Last sequence update)			
DT	01-DEC-2001	(Tremblrel. 19, Last annotation update)			
DE	COMPLEMENT FACTOR C2 (COMPLEMENT COMPONENT 2) (WITHIN H-2S).				
GN	C2.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=129;				
RA	Rowen L., Madan A., Qin S., Hall J., Dahl T., James R., Dickhoff R.,				
RA	Schaffer T., Ratcliffe A., Abbasi N., Loretz C., Lasky S., Hood L.;				
RT	"Sequence of the mouse MHC class III region."				
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE OF 291-760 FROM N.A.				
RA	Rowen L., Qin S., Lasky S.R., Loretz C., Dors M., Mahairas G.,				
RA	Hood L.;				
RL	Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=BREAST TUMOR;				
RA	Straussberg R.;				
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.				
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE				
CC	TRYPSIN FAMILY.				
DR	EMBL;	AF109906;	AAC84162.1;		
DR	EMBL;	AF049850;	AAC05284.1;		
DR	EMBL;	BC011086;	AAH11086.1;		
DR	HSSP;	P00761;	1EPT.		
DR	MGD;	MGI:88226;	C2.		
DR	InterPro;	IPR001314;	Chymotrypsin.		
DR	InterPro;	IPR000436;	Sushi_SCR_CCP.		
DR	InterPro;	IPR001254;	Trypsin.		
DR	InterPro;	IPR002035;	VFPA.		
DR	Pfam;	PF00089;	trypsin;	2.	
DR	Pfam;	PF00092;	vwa;	1.	
DR	PRINTS;	PR00722;	CHYMOTRYPSIN.		
DR	PRINTS;	PR00453;	VWFA.		
DR	SMART;	SM00032;	CCP;	2.	
DR	SMART;	SM00020;	Tryp_Spc;	1.	



```

Query Match      6.9%; Score 132.5; DB 5; Length 765;
Best Local Similarity 25.1%; Pred. No. 0.0063;
Matches 50; Conservative 35; Mismatches 89; Indels 25; Gaps

QY 39 CYGGFDLYFLDKSGSVLH-HMNEIYFVFEOLAHKFIS-----POLRMSFIVFSTRGTT 91
Db 72 CTSLQDICFLVDSGSGIGENHYEE---VKQFLHAFLSKLPIGNDEVNTSLVLFSTTVHP 127

QY 92 LMKL-----TEDREIQIRQGLEELQKVLPGDGYTMHEGFERASEQIY-YENRQGYRTASVII 146
Db 128 HWSLRANNASDKETAMQ--DVLPTIPYHGTTNTAAGLQTCNQMLFDYPRERQTVPEKLVI 185

QY 147 ALTDGELHEDLPFYSREANRSRDGLAIYVCVGKQDF-NETQLARLADSKDHVFP----V 201
Db 186 AMTDGSDSD--EFTYNEAKVIRERGIITVLSVGMVYVNHNECRSMCGCRNDSPPCLYL 243

QY 202 NDGFQALQGIHLSILKSC 220
Db 244 QTWSQLLPSISPILKEVC 262

RESULT 11
O43853 PRELIMINARY; PRT; 517 AA.
ID O43853 PRELIMINARY; PRT; 517 AA.
AC O43853;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TYPE XII COLLAGEN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=CORNEA;
RX MEDLINE=980033048; PubMed=9344363;
RA Wessel H., Anderson S., Fite D., Halvas E., Hempel J., SundarRaj N.;
RT "Type XII collagen contributes to diversities in human corneal and
RL limbal extracellular matrices.";
RL Invest. Ophthalmol. Vis. Sci. 38:2408-2422(1997).
DR EMBL; U68139; AAC01506.1; .
DR HSSP; P17301; LAOX.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00041; fn3; 3.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00060; FN3; 3.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWFA; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 517
FT SEQUENCE 517 AA; 56727 MW; 9B6972F44A1BD88F CRC64;

Query Match      6.9%; Score 131.5; DB 4; Length 517;
Best Local Similarity 29.6%; Pred. No. 0.0045;
Matches 61; Conservative 39; Mismatches 77; Indels 29; Gaps

QY 39 CYGGFDL---YFLDKSGSV-LHHMNEIYFVFEOLAHKF-ISP-QLRMSFIVFST---RG 89
Db 133 CSRGVDIKADMFLVDGSGYIGIANVVKVRAFLFVLVKFSFISPNRQVLSUVQYSRDPHTE 192

QY 90 TTLMLKLTREQIRQGLEELQKVLPL--GGDTYMHGFERASEQIYYENRQGYRT--ASVI 145
Db 193 FTLLKKTKVEDII-----EAINTFPYRGSGTNTGKAMTYVREKLIYFVPSK-GSRSNYPKVM 246

QY 146 IALTLDGELHEDLPFYSREANRSRDGLAIYVCVGKDFNETQLARIAD---SKOHVFPVND 203
Db 247 ILITDGR-SSDAF---RDPAIKLRNSDVEIFAVGVKDAVDALEAATASPPAEHTVFTEVD 302

```

us-09-970-076-2.rspt

Fri Aug 9 10:57:01 2002

```

QY 204 GFOALQGIHHSILKKSCIEI---LAA 226
DB 303 -FDAFQISFELTQISCLRIEQELAA 327

RESULT 12
Q9UGC3 PRELIMINARY; PRT; 660 AA.
ID Q9UGC3
AC Q9UGC3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE D234P15.1 (COLLAGEN, TYPE XII, ALPHA 1) (FRAGMENT).
GN COL12A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith M.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL080250; CAB65984.1; -.
DR HSP; P11215; IBHQ.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00092; vwa; 2.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00327; VWA; 2.
DR PROSITE; PS50234; VWFA; 2.
KW Collagen.
KW COLLAGEN.
FT NON_TER 660
FT SEQUENCE 660 AA; 72523 MW; 3E3A10A285ECA51 CRC64;

Query Match 6.8%; Score 131; DB 4; Length 660;
Best Local Similarity 29.8%; Pred. No. 0.007;
Matches 59; Conservative 38; Mismatches 75; Indels 26; Gaps 13;

QY 44 DLYFTLDKSGV-LHHWNEIYFVQLAHKF-ISP-QLRMSFIVFST---RGTTLMKLTE 97
DB 469 DIVFLVDGYSIGIANFVKVRAFLVVKVSEISPNRVQISLVQSRDPHTFTLKPKTK 528
QY 98 DREQIRQGLEELQKVLPGGDTYMHGFEFASQIYYENRQGYR--ASVITIALTDGEL 153
DB 529 VEDII-----EAIPTFPYEGSGTNGKAMTYVREKIFVPSK-GSRSNVPKVMILITDCK- 581
QY 154 HEDLFYSERANRSDLGAIYVCVGVKDFNETQLARIAD--SKDHVPVNDGFOALOGI 211
DB 582 SSDAF---RDPALKRNSDVEIFAVGVKDAVRSELEAIASPPAETHVETVED-FDAFORI 637
QY 212 IHSILKSCIEI---LAA 226
DB 638 SFELTQISCLRIEQELAA 655

RESULT 13
Q28984 PRELIMINARY; PRT; 920 AA.
ID Q28984
AC Q28984;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CD1IB (FRAGMENT).
GN CD1IB.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.

Lee J.-K., Schook L.B., Rutherford M.S.;
Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
EMBL; U40072; AAB16869.1; -.
DR HSP; P11215; IABX.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF01839; FG-GAP; 4.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 4.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWFA; 1.
FT NON_TER 920
FT SEQUENCE 920 AA; 102440 MW; E96CC51E350DD5AC CRC64;

Query Match 6.8%; Score 130.5; DB 6; Length 920;
Best Local Similarity 26.4%; Pred. No. 0.012;
Matches 63; Conservative 41; Mismatches 72; Indels 63; Gaps 15;

QY 44 DLYFTLDKSGV-LHHWNEIYFVQLAHKFISP---QLRMSFIVFSTRTTLMKLTE-- 98
DB 17 DIAFLIDSGSI---NRLDF---QRMKEFVSTVMGQPKSKTLFA-----LMQYSEDFY 64
QY 99 -----REQIRQGLEELQKVLPGGDTYMHGFEFASQIYYENRQGYR- 140
DB 65 THFTFNDFKRNPSPKLVRPIRQLL-----ORTHTATGIRKVVRELFH-SKSGARE 114
QY 141 -TASVITIALTDGELHEDLFYSE--REANRSDLGAIYVCVGVKDFN-----ETQLARIA 192
DB 115 NALKILVITDGKFGDPLGYEDVPEADRK---GVIRYVIGVGDAFNWSKRSREELNTIA 171
QY 193 DSK--DHVFPVNDGFOALOGIHSILKKSCIEILAAEPSTICAGESFOVYVNGNGFRHA 249
DB 172 SKPCGDHVFQVNN-PEAVKTIQNLQKLT-----FAIEGTGTGTSTSPFCEMSQEGFSNA 225

RESULT 14
Q9UA13 PRELIMINARY; PRT; 3767 AA.
ID Q9UA13
AC Q9UA13; Q21340;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TRANSMEMBRANE CELL ADHESION RECEPTOR MUA-3 PRECURSOR (K08E5.3 PROTEIN) (FRAGMENT).
GN MUA-3 OR K08E5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N2;
RA Lu Z., Vogel B., Hedgecock E.;
RA "mua-3 mRNA Splicing Pattern Revealed.";
FT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RL [2]
RN SEQUENCE FROM N.A.
RP Kershaw J.K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; AF139060; AAD29428.1; -.
DR EMBL; Z30974; CAA83226.2; -.
DR EMBL; Z30423; CAA83226.2; JOINED.

```



us-09-970-076-2.rspt

Fri Aug 9 10:57:01 2002

\_\_\_\_\_



```

RESULT      3
RWHU1C
cell surface glycoprotein CD11c precursor - human
N;Alternate names: leukocyte adhesion receptor p150,95 alpha chain
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 22-Jun-1999
C;Accession: A36584; A35543; S00864
R;Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.
J. Biol. Chem. 265, 12750-12751, 1990
A;Reference number: A36584
A;Contents: erratum
A;Accession: A36584
A;Molecule type: DNA
A;Residues: 1-1163 <COR>
A;Note: this revision to the sequence from reference A35543 includes the carboxyl end
R;Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.
J. Biol. Chem. 265, 2782-2788, 1990
A;Title: Genomic structure of an integrin alpha subunit, the leukocyte p150,95 molecule
A;Reference number: A35543; MUID:90153906
A;Accession: A35543
A;Molecule type: DNA
A;Residues: 1-834 <CO2>
A;Note: this sequence has been revised in reference A36584
A;Accession: A36584; Watson, P.S.; Springer, T.A.

```







R:Gerecke, D.R.; Foley, J.W.; Castagnola, P.; Gennari, M.; Dublet, B.; Cancedda, R.; Llin  
J. Biol. Chem. 268, 12177-12184, 1993  
A:Title: Type XIV collagen is encoded by alternative transcripts with distinct 5' region  
ns.

A:Reference number: A45974; MUID:93280195

A:Accession: A45974

A>Status: preliminary

A:Molecule type: mRNA; protein

A:Residues: 1-1747 <GER>

A:Experimental source: embryo skin

A>Note: sequence inconsistent with the nucleotide translation

A>Note: sequence extracted from NCBI backbone (NCBI:133364, NCBI:133365)

R:Apte, S.S.

submitted to the EMBL Data Library, March 1992

A:Reference number: S30085

A:Accession: S30085

A:Molecule type: mRNA

A:Residues: 1472-1660 <APT>

A:CROSS-references: EMBL:X65122; NID:962871; PIDN:CAA46238.1; PID:g938175

R:Trueb, J.; Trueb, B.

Eur. J. Biochem. 207, 549-557, 1992

A:Title: Type XIV collagen is a variant of undulin.

A:Reference number: S22916; MUID:92339443

A:Accession: S22916

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 286-494, 'Q', 496-834, 'A', 836-1119, 'KL', 1122-1402, 1409-1439 <TRU>

R:Gordon, M.K.; Castagnola, P.; Dublet, B.; Linsenmayer, T.F.; van der Rest, M.; Mayne,

Eur. J. Biochem. 201, 333-338, 1991

A:Title: Cloning of a cDNA for a new member of the class of fibril-associated collagens

A:Reference number: S17035; MUID:92037585

A:Accession: S17035

A:Molecule type: mRNA

A:Residues: 1472-1659 <GOR1>

A:Accession: S20833

A:Molecule type: protein

A:Residues: 1551-1570; 1593-1599; 1639-1667 <GOR2>

C:Superfamily: collagen alpha 1(XIV) chain; fibronectin type III repeat homology; von W

C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trim

F:40-204/Domain: von Willebrand factor type A repeat homology <VWAL>

F:236-317/Domain: fibronectin type III repeat homology <FN3A>

F:326-409/Domain: fibronectin type III repeat homology <FN3B>

F:418-498/Domain: fibronectin type III repeat homology <FN3C>

F:507-591/Domain: fibronectin type III repeat homology <FN3D>

F:625-707/Domain: fibronectin type III repeat homology <FN3E>

F:716-798/Domain: fibronectin type III repeat homology <FN3F>

F:808-893/Domain: fibronectin type III repeat homology <FN3G>

F:924-1089/Domain: von Willebrand factor type A repeat homology <VWA2>

F:1111-1352/Domain: non-collagenous NC4 #status predicted <NC4>

F:1511-1553/Domain: non-collagenous NC2 #status predicted <NC2>

F:1554-1659/Domain: triple helical domain COL1 #status predicted <COL1>

Query Match 7.2%; Score 137; DB 1; Length 1747;  
Best Local Similarity 25.4%; Pred. No. 0.016;  
Matches 63; Conservative 45; Mismatches 114; Indels 26; Gaps 12;

QY 44 DLYFTLDKSGV-LHHWNIYFVEQL--AHKFISPO-LRMSFIVFSTRTLMKLT--E 97

Db 926 DLVFLVDGWSIGDDNFNKKISFLSTVGALDKIGDGTQVAIIQFSDPRTEFKLNAYK 985

QY 98 DREQIRGLEELQKLVPGDGYMHGFEFRASEQIYYENRQGYRTA--SVIATLDGELHE 155

Db 986 TKETLEALQI--AYKGNKTKGAIKHARE-VLFTGEAGNRKGIKPVLVVITDGRSQD 1042

QY 156 DLFFYSEREANRDL-GAIVYCVGVKDFNETQLARIAD--SKDHVPVNDGFMALQGI 212

Db 1043 DY-----NKVSREMOLDGFSFAIGVADADYSELNIGSKPSRHHVFFVDD-FDAFTKIE 1096

QY 213 HSILAKSCIEILAAEPSITCAGESQVVVVGNGFRHARNDVRLVCSFKINDSVTLNKP 272

Db 1097 DELITVFCETASATCPLVFKDGDNEFA-----GFKMMEMFGLVEKFEAISDGVSMPEPT 1150

QY 273 SVEDTYLL 280  
Db 1151 NVVPCYRL 1158

RESULT 8

C2MS

Classical-complement-pathway C3/C5 convertase (EC 3.4.21.43) C2 component precursor -  
N:Alternate names: C3 convertase; C5 convertase; complement C2

C:Species: Mus musculus (house mouse)

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 19-May-2000

C:Accession: A38876; B36593; I54429

R:Ishikawa, N.; Nonaka, M.; Wetsel, R.A.; Colten, H.R.

submitted to GenBank, January 1991

A:Reference number: A38875

A:Accession: A38876

A:Molecule type: DNA

A:Residues: 1-760 <IS2>

A:CROSS-references: GB:M57891; GB:J05661; NID:g192436; PIDN:AAA63294.1; PID:g192437

R:Ishikawa, N.; Nonaka, M.; Wetsel, R.A.; Colten, H.R.

J. Biol. Chem. 265, 19040-19046, 1990

A:Title: Murine complement C2 and factor B genomic and cDNA cloning reveals different

A:Reference number: A36593; MUID:91035430

A:Accession: B36593

A:Molecule type: mRNA

A:Residues: 1-760 <ISH>

A:CROSS-references: EMBL:M57891; NID:g192436; PIDN:AAA63294.1; PID:g192437

R:Palus, A.; Wakeland, E.K.; McConnell, T.J.; Gitlin, J.; Whitehead, A.S.; Colten, H.

Immunogenetics 25, 290-298, 1987

A:Title: DNA polymorphism of MHC III genes in inbred and wild mouse strains.

A:Reference number: I54429; MUID:87192938

A:Accession: I54429

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 660-677, 'R', 679, 681-723, 'G', 725 <RES>

A:CROSS-references: GB:MI6271; NID:g199289; PIDN:AAA39562.1; PID:g199290

C:Genetics:

C:Introns: 16/1; 91/1; 153/1; 212/1; 245/3; 290/3; 337/1; 384/1; 414/1; 461/1; 492/3;

C:Complex: The proenzyme forms a complex with C4a and is activated by cleavage into C

C:Function:

A:Description: cleaves complement C3 and complement C5 alpha chains

A:Pathway: complement classical pathway

C:Superfamily: complement C2; complement factor H repeat homology; trypsin homology;

C:Keywords: alternative splicing; complement classical pathway; duplication; glycopro

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-250/Product: complement C2b fragment #status predicted <C2B>

F:22-89/Domain: complement factor H repeat homology <FH1>

F:94-149/Domain: complement factor H repeat homology <FH2>

F:156-210/Domain: complement factor H repeat homology <FH3>

F:251-760/Product: complement C2a fragment long form #status predicted <C2A>

F:259-449/Domain: von Willebrand factor type A repeat homology <VFA>

F:478-747/Domain: trypsin homology #status atypical <TRY>

F:22-62,49-89,94-136,122-149,156-197,182-210,470-590,499-515,593-609,647-674,685-715/

F:27,117,297,340,474,478,663/Binding site: carbohydrate (Asn) (covalent) #status pred

F:514,570,689/Active site: His, Asp, Ser #status predicted

Query Match 6.8%; Score 131; DB 1; Length 760;  
Best Local Similarity 23.2%; Pred. No. 0.016;  
Matches 74; Conservative 56; Mismatches 115; Indels 74; Gaps 16;

QY 41 GGFDLYFLDKSGV-LHHWNIY-VFVEQLAHKFISPOLR--MSFIVFSTRTLMKLT 97

Db 258 GHNLNLYLLDASQSTKEDFDIFKKSAELMVERISFEVNVITVAITFASQPKTMSILS 317

QY 98 DREQIRGLEELQKLVPGDGYMHGFEFRASEQIYYE-----NROGYRTAS- 143

Db 318 ERSQ-----DVTEVITSLDSASYKDHEATCATNYEVLIRVYSMMQTMQMDRLGMEISAW 371

QY 144 -----VITALTDGELHEDLFFYSEREANRDLGAI-----VYCVGV-----KD 182

Db 372 KEIRTHITLLTDGK--SNMGDSPKKAIVTRIRELLSIEQNRRDYLDTIAYIGVGKLDVDWKE 429

QY	183	FNETOLARIADSKDHVFPVNDGFOALQGIHSHILKSCIEILAAEPSTICAGESFOVVVR	242
Db	430	LNE--LGSKKDGERHAFILODA-KALQIFHEHMLDYSKL-----TDTICG-----V	472
QY	243	GNGFRIHARNVDRV--LCSFKINDSVILNEKPFVSVEDTYLLCPAPILKEVGMK--AALQVYS	298
Db	473	GNMSANASQDERTPWQVTFKPKSKETCOGS--LISDQWILTAACHFDIQMEDHHLWRVN	530
QY	299	MND-----GLSFTSSSVII	312
Db	531	VGDPSTQHGKEFLVEDVII	549
RESULT	9		
S31212			
Collagen alpha 1(XIV) chain precursor, short form - chicken			
C:Species: Gallus gallus (chicken)			
C:Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 13-Aug-1999			
C:Accession: S31212			
R:Waelchli, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.			
Eur. J. Biochem. 212, 483-490, 1993			
A:Title: Complete primary structure of chicken collagen XIV.			
A:Reference number: S31211; MUID:93185668			
A:Accession: S31212			
A:Status: nucleic acid sequence not shown; translation not shown			
A:Molecule type: mRNA			
A:Residues: 1-1857 <WAE>			
A:Cross-references: EMBL:X70792; NID:g288874; PIDN:CAA50063.1; PID:g288875			
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993			
C:Genetics:			
A:Gene: Coll14A1			
C:Superfamily: collagen alpha 1(XIV) chain; fibronectin type III repeat homology; von Willebrand factor type A repeat homology			
C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer			
F:1-28/Domain: signal sequence #status predicted <Sig>			
F:29-1857/Product: collagen alpha 1(XIV) chain, short form #status predicted <MAT>			
F:29-110/Domain: fibronectin type III repeat homology <FN3A>			
F:156-320/Domain: von Willebrand factor type A repeat homology <VWA1>			
F:352-433/Domain: fibronectin type III repeat homology <FN3B>			
F:442-525/Domain: fibronectin type III repeat homology <FN3C>			
F:534-614/Domain: fibronectin type III repeat homology <FN3D>			
F:623-707/Domain: fibronectin type III repeat homology <FN3E>			
F:741-823/Domain: fibronectin type III repeat homology <FN3F>			
F:832-914/Domain: fibronectin type III repeat homology <FN3G>			
F:922-1009/Domain: fibronectin type III repeat homology <FN3H>			
F:1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>			
Query Match	6.8%	Score 130; DB 2; Length 1857;	
Best Local Similarity	25.0%	Pred. No. 0.065;	
Matches	62; Conservative	44; Mismatches 116; Indels 26; Gaps 12;	
QY	44	DLIFLDKSGSV-LHHWNEIYFVEQL--AHKFISPO-LRMSFVSTRTTLMKLT--E	97
Db	1042	DLVFLVGDGWSIGDDNFNKLISFLYSTVGALDKIGDGTQVAILIQFSDDPRTFEKLNAYK	1101
QY	98	DREIQRGLEELQKVLPGDGTMYHGEFPERASEQIYYENRQGYRTA--SVIIALTDGELHE	155
Db	1102	TKETLLEAIQOI--AYKGGNTKTGKAIKHARE--VLFTGEAGMRKGIPKVLVITDGRSQD	1158
QY	156	DLFFYSREANRSDL-GAIVYCVGVKDFNETQLARIAD--SKDHVPVNDGFOALQGI	212
Db	1159	DV-----NKVSRMQLDGFSSFAIGVADADYSELVNIWIGSKPSERHVFVDD-FAFAFKIE	1212
QY	213	HSILKKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVDRVLCSPKINDSVTLNKPFF	272
Db	1213	DELITFCVETASATCPLVFKDQKLA-----GFKMMEMFGLVEKEFSIDGVSMPEPTGF	1266
QY	273	SVEDTYLL	280
Db	1267	NVYPCYRL	1274
RESULT	11		
S42373			
Hypothetical protein T20G5.3 - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C:Date: 07-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999			
C:Accession: S42373			
R:Smith, A.			
submitted to the EMBL Data Library, March 1994			



A:Accession: S34075  
A:Molecule type: mRNA  
A:Residues: 1-764 <MEJ>  
A:Cross-references: EMBL:X72875; MID:g297568; PIDN:CRA51389.1; PID:g297569  
R:Woods, D.E.; Markham, A.F.; Ricker, A.T.; Goldberger, G.; Colten, H.R.  
Proc. Natl. Acad. Sci. U.S.A. 79, 5661-5665, 1982  
A:Title: Isolation of cDNA clones for the human complement protein factor B, a class III  
A:Reference number: A44622; MUID:83039428  
A:Accession: A44622  
A:Molecule type: mRNA  
A:Residues: 467-546; 550-595; 752-764 <WOO>  
A:Cross-references: GB:J00185; GB:J00186  
A:Note: The authors translated the codon TAC at 519 as Thr; the nucleic acid translation  
R:Mole, J.E.; Anderson, J.K.; Davison, E.A.; Woods, D.E.  
J. Biol. Chem. 259, 3407-3412, 1984  
A:Title: Complete primary structure for the zymogen of human complement factor B.  
A:Reference number: A20751; MUID:84161997  
A:Accession: A00934  
A:Molecule type: protein; mRNA  
A:Residues: 26-764 <MOL>  
A:Cross-references: GB:K01566  
A:Note: nucleic acid translation differs from the sequence shown in having 300-Leu, 328-  
A:Note: 736-Ser was also found  
A:Note: glycosylation sites were determined  
R:Christie, D.L.; Gagnon, J.  
Biochem. J. 209, 61-70, 1983  
A:Title: Amino acid sequence of the Bb fragment from complement factor B. Sequence of the  
A:Reference number: A19188; MUID:83204002  
A:Contents: the final paper in a series documenting the sequence, glycosylation site, and  
A:Accession: A19188  
A:Molecule type: protein  
A:Residues: 260-296, 298-764 <CHR>  
R:Campbell, R.D.; Porter, R.R.  
Proc. Natl. Acad. Sci. U.S.A. 80, 4464-4468, 1983  
A:Title: Molecular cloning and characterization of the gene coding for human complement  
A:Reference number: A19947; MUID:83273641  
A:Accession: A19947  
A:Molecule type: DNA  
A:Residues: 346-764 <CAM>  
A:Cross-references: GB:J00125  
A:Accession: B19947  
A:Molecule type: mRNA  
A:Residues: 339-509 <CAL>  
A:Cross-references: GB:J00126; MID:g187723; PIDN:AAA36226.1; PID:g553536  
R:Wu, L.; Morley, B.J.; Campbell, R.D.  
Cell 48, 331-342, 1987  
A:Title: Cell-specific expression of the human complement protein factor B gene: evidence  
A:Reference number: A25971; MUID:87102880  
A:Accession: B25971  
A:Molecule type: DNA  
A:Residues: 1-99 <MUL>  
A:Cross-references: GB:M15082; MID:g187699; PIDN:AAA59625.1; PID:g553534  
R:Niemann, M.A.; Brown, A.S.; Miller, E.J.  
Biochem. J. 274, 473-480, 1991  
A:Title: The principal site of glycation of human complement Factor B.  
A:Reference number: S14339; MUID:91174758  
A:Accession: S14339  
A:Molecule type: protein  
A:Residues: 270-329 <NIE>  
A:Note: binding site for carbohydrate to lysine under artificial conditions  
R:Morley, B.J.; Campbell, R.D.  
EMBO J. 3, 153-157, 1984  
A:Title: Internal homologies of the Ba fragment from human complement component factor B  
A:Reference number: A44628; MUID:84158524  
A:Accession: A44628  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 16-225, 227-259 <MOR>  
R:Schwaible, W.; Luttig, B.; Sokolowski, T.; Estaller, C.; Weiss, E.H.; Meyer zum Busche  
Immunobiology 186, 221-232, 1993  
A:Title: Human complement factor B: functional properties of a recombinant zymogen of the  
A:Reference number: I54409; MUID:94041399  
A:Accession: I54409

A:Status: translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-764 <RES>  
A:Cross-references: GB:S67310; MID:g452937; PIDN:AAD13989.1; PID:g4261689  
R:Horuchi, T.; Kim, S.; Matsumoto, M.; Watanabe, I.; Fujita, S.; Volanakis, J.E.  
Mol. Immunol. 30, 1587-1592, 1993  
A:Title: Human complement factor B: cDNA cloning, nucleotide sequencing, phenotypic c  
A:Reference number: I57824; MUID:94067177  
A:Accession: I57824  
A:Status: translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-31, 33-764 <RE2>  
A:Cross-references: GB:I15702; MID:g291921; PIDN:AAA16820.1; PID:g291922  
C:Comment: 292-Cys has a free sulfhydryl.  
C:Genetics:  
A:Gene: GDB:BF  
A:Cross-references: GDB:I19726; OMIM:138470  
A:Map position: 6p21.3-6p21.3  
A:Introns: 21/3; 99/3; 346/1; 390/1; 424/1; 470/1; 502/3; 542/1; 593/2; 619/1; 652/3;  
A:Note: the list of introns may be incomplete  
A:Note: gene is located in the major histocompatibility complex, class III region  
C:Complex: complement factor B initially forms an inactive complex with complement fa  
ment factor C3b forming active C3/C5 convertase; Ba is released  
C:Function:  
A:Description: Bb is a serine proteinase; C3/C5 convertase cleaves complement C3 alph  
a1  
A:Pathway: complement alternate pathway  
C:Superfamily: complement C2; complement factor H repeat homology; trypsin homology;  
C:Keywords: acute phase; complement alternate pathway; duplication; glycoprotein; hyd  
F:1-25/Domain: signal sequence #status predicted <SIG>  
F:26-764/Product: complement factor B #status experimental <BAF>  
F:26-259/Product: complement factor Ba fragment #status experimental <BAF>  
F:37-98/Domain: complement factor H repeat homology <FH1>  
F:103-158/Domain: complement factor H repeat homology <FH2>  
F:165-218/Domain: complement factor H repeat homology <FH3>  
F:260-764/Product: C3/C5 convertase Bb fragment #status experimental <BBF>  
F:268-458/Domain: von Willebrand factor type A repeat homology <VFA>  
F:482-752/Domain: trypsin homology #status atypical <TRY>  
F:37-76, 62-98, 103-145, 131-158, 165-205, 191-218, 478-596, 511-527, 599-615, 656-682, 695-725  
F:122, 142, 285, 378/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:259, 260/Cleavage site: Arg-Lys (complement factor D) #status experimental  
F:526, 576, 699/Active site: His, Asp, Ser #status experimental

Query Match	5.98;	Score 113.5;	DB 1;	Length 764;
Best Local Similarity	19.68;	Pred. No. 0.43;	Indels 99;	Gaps 17;
Matches 66;	Conservative	63;	Mismatches	109;
QY 28	QCGRRDGG-GPA-----	CYGGFDLYFLDKSGSV-----	IHHWNIYYFVEQLAH 71	
DB 243	EGVDREDGHPGEQQRKRVLDPSGSMNLYLVLDGSDSIGASNFTGAKKCLVNLIERVAS 302			
QY 72	KFISPLRMSFTVFTGRGTTLMKLFTE----	DREQIRQGLEEL----	QKVLPGGDYMHGEG 123	
DB 303	YGVKVP-RYGLVYATYPRKIWVKVSEADSSNADWYTKOLNEINYEDHKLKSGTNT-----	355		
QY 124	FRASEFQIYYENR-----	QGY-RTASVITALTDTG-----	ELHEDLFFYSE 162	
DB 356	-KKAQAVTSMMSWFDVPPPEGNRTRHVIILMTDGLHNMGGDPITVDEIRDLLYIGKD 414			
QY 163	REANRSRLDGAIVYCVG--VKDFNETQIARIADSKDHFVPPVNDGFOALQGIHLSILKSKS 220			
DB 415	RKNPRELDYDVYVFGVGPLVNOVNINALASKKNEQHFVKVD-MENLEDVYFQIMIDES- 472			
QY 221	TEILAAEPSTICAGESFQVVRGNGFRHARNVDRLCSEKINDSVTLNKPFSVEDTYLL 280			
DB 473	-----QSLSLC-----	GMVWEHRKGTD-----	YHKQPQWAKISV-- 501	
QY 281	CPAPILKEVGMKAAALQVSMNDGLSFSSSVIITTHC 317			
DB 502	-----IRPSKGHESCMG-----	AVVSEFVLTAHC 527		

## RESULT 15

A54849  
Collagen alpha 1(VII) chain precursor - human  
N:Alternate names: procollagen alpha 1(VII) chain  
C:Species: Homo sapiens (man)  
C:Date: 04-Nov-1994 #sequence\_revision 04-Nov-1994 #text\_change 20-Sep-1999  
C:Accession: A54849; PH0844; S16316; I56328; A30296; I84686  
R:Christiano, A.M.; Greenspan, D.S.; Lee, S.; Uitto, J.  
J. Biol. Chem. 269, 20256-20262, 1994  
A:Title: Cloning of human type VII collagen. Complete primary sequence of the alpha1(VII)  
A:Reference number: A54849; MUID:94327588  
A:Accession: A54849  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-2944 <CHR>  
A:Cross-references: GB:L02870; NID:9987124; PIDN:AAA75438.1; PID:9987125  
R:Tanaka, T.; Takahashi, K.; Furukawa, F.; Imamura, S.  
Biochem. Biophys. Res. Commun. 183, 958-963, 1992  
A:Title: Molecular cloning and characterization of type VII collagen cDNA.  
A:Reference number: PH0844; MUID:92231902  
A:Accession: PH0844  
A:Molecule type: mRNA  
A:Residues: 'E'F, 340-475, 'RALSTASHSTLCWTRWHPNCGSHWTRACPCNRPASHRAARAG', 524-528, 'C',  
A:Cross-references: DDBJ:D1152; DDBJ:D13694; NID:9453698; PIDN:BAA02853.1; PID:9453699  
A:Experimental source: Keratinocyte  
A:Note: the authors translated the codon ACC for residues 394 and 397 as Tyr  
R:Parente, M.G.; Chung, L.C.; Ryyanaenen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Mat  
Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991  
A:Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.  
A:Reference number: S16316; MUID:91334380  
A:Accession: S16316  
A:Molecule type: mRNA  
A:Residues: 815-892, 'E', 894-1439 <PAR>  
A:Cross-references: GB:M65158; GB:S49017; NID:g180914; PIDN:AAA96439.1; PID:g180915  
A:Experimental source: Keratinocyte  
R:Gammon, W.R.; Abernethy, M.L.; Padilla, K.M.; Prisayanah, P.S.; Cook, M.E.; Wright, J.;  
J. Invest. Dermatol. 99, 691-696, 1992  
A:Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion prot  
A:Reference number: I56328; MUID:93107742  
A:Accession: I56328  
A:Status: translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 'E'F, 372-517, 'DV', 520-540, 'W', 542-1255 <RES>  
A:Cross-references: GB:S51236; NID:g262308; PIDN:AAB24637.1; PID:g262309  
R:Seitzer, J.L.; Eisen, A.Z.; Bauer, E.A.; Morris, N.P.; Glanville, R.W.; Burgeson, R.E.  
J. Biol. Chem. 264, 3822-3826, 1989  
A:Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagena  
A:Reference number: A30296; MUID:89139437  
A:Accession: A30296  
A:Molecule type: protein  
A:Residues: 'A', 1240-1246, 'G', 1248-1250, 'XE', 1253-1255, 'Q', 1257, 'E', 2032, 'C', 2034-2041,  
A:Note: two reported peptides cannot be reliably located  
R:Greenspan, D.S.  
Hum. Mol. Genet. 2, 273-278, 1993  
A:Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous  
A:Reference number: I48103; MUID:93271985  
A:Accession: I48103  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 2395-2871, 'S', 2873-2944 <RE2>  
A:Cross-references: GB:L06862; NID:g388713; PIDN:AAA89196.1; PID:g388714  
R:Christiano, A.M.; Ryyanaenen, M.; Uitto, J.  
Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994  
A:Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --> Ser subs  
A:Reference number: A55255; MUID:94224777  
A:Contents: annotation  
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C  
ed and subsequently O-glycosylated.  
C:Genetics:  
A:Gene: GDB:COL7A1; EBR1; EBD1; EB  
A:Cross-references: GDB:128750; OMIM:120120  
A:Map position: 3p21.3-3p21.3  
A:Note: defects in this gene can result in dominant and recessive dystrophic epidermolys

A:Note: there are 118 introns  
C:Complex: type VII collagen is probably a homotrimer  
C:Function:  
A:Description: structural component of extracellular polymer associated with anchoring  
C:Superfamily: unassigned collagens; animal kunitz-type proteinase inhibitor homology  
C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxypr  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>  
F:17-1253/Domain: amino-terminal nonhelical #status predicted <NC1>  
F:36-201/Domain: von Willebrand factor type A repeat homology <VWA1>  
F:231-318/Domain: fibronectin type III repeat homology <FN1>  
F:327-413/Domain: fibronectin type III repeat homology <FN2>  
F:414-502/Domain: fibronectin type III repeat homology <FN3>  
F:508-593/Domain: fibronectin type III repeat homology <FN4>  
F:598-683/Domain: fibronectin type III repeat homology <FN5>  
F:686-771/Domain: fibronectin type III repeat homology <FN6>  
F:776-862/Domain: fibronectin type III repeat homology <FN7>  
F:864-952/Domain: fibronectin type III repeat homology <FN8>  
F:954-1045/Domain: fibronectin type III repeat homology <FN9>  
F:1052-1219/Domain: von Willebrand factor type A repeat homology <VWA2>  
F:1170-1172/Region: cell attachment (R-G-D) motif  
F:1189-1253/Region: cysteine/proline-rich  
F:1254-2783/Region: interrupted helical  
F:1334-1336/Region: cell attachment (R-G-D) motif  
F:2008-2010/Region: cell attachment (R-G-D) motif  
F:2533-2555/Region: cell attachment (R-G-D) motif  
F:2784-2944/Domain: carboxyl-terminal nonhelical #status predicted <NC2>  
F:2876-2929/Domain: animal kunitz-type proteinase inhibitor homology <BPI>  
F:337,786,1109/Binding site: carbohydurate (Asn) (covalent) #status predicted  
F:2167,2176,2185,2188,2664,2667,2673/Modified site: 4-hydroxyproline (Pro) #status ex  
F:2625,2631/Binding site: 5-hydroxylysine (Lys) #status experimental  
F:2625,2631/Binding site: carbohydurate (Lys) (covalent) #status experimental  
F:2634,2802,2804/Disulfide bonds: interchain #status predicted  
Query Match 5.8%; Score 111.5; DB 2; Length 2944;  
Best Local Similarity 24.1%; Pred. No. 3.9;  
Matches 52; Conservative 42; Mismatches 85; Indels 37; Gaps 11:  
Qy 44 DLYFLDKSGSV-LHHWNEIYFVEQLAHKF---ISPO-LRMSFTVFSTRGTT---LMKL 95  
Db 38 DIVFLDGSSTIGRSNFRVRSFLEGLVLPFGSAASAGVRFATVQYSDDPRTFGLDAL 97  
Qy 96 TEDREIQIRGLEELQKVLPGDVTYMHGFERASEIQIYYENRQGYRTASVIALTDGELHE 155  
Db 98 GSGGDVIR-ATRELS--YKGNTRTGAALHVADHVELPQLARPCVKVILITDGK-SQ 153  
Qy 156 DLFYVSEREANRDLGAIVYCVGVKDFNETQLARIAD--SKDHVPVNDGFQALOGIIH 213  
Db 154 DLV---DTAAORLKGQVKLEAVGIKNADPELKRVASQPTSDFFFFVND-FSILRTLIP 209  
Qy 214 SILKSCSI-----ETLAAEPST 230  
Db 210 LVSRVCTTAGGVVTRPPDDSTSPAPROLVSEPS 245  
Search completed: August 9, 2002, 10:34:12  
Job time: 251 sec

us-09-970-076-2.rpr

Fri Aug 9 10:57:00 2002

-----

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1090.8	98.8	98.8	5540	1	PCT-US02-08253-176	Sequence 176, App
2	1090.8	98.8	98.8	5540	1	PCT-US02-08253-231	Sequence 231, App
3	960.8	87.0	1674	7	US-10-038-307-17	Sequence 17, Appl	
4	952.4	86.3	1650	7	US-10-038-307-13	Sequence 13, Appl	
5	952.4	86.3	1650	7	US-10-038-307-15	Sequence 15, Appl	
6	950.8	86.1	1650	7	US-10-038-307-9	Sequence 9, Appl	
7	949.4	86.0	1056	7	US-10-038-307-23	Sequence 23, Appl	
8	949.4	86.0	1713	7	US-10-038-307-19	Sequence 19, Appl	
9	949.4	86.0	2272	7	US-10-038-307-1	Sequence 1, Appl	
10	949	86.0	1008	7	US-10-038-307-25	Sequence 25, Appl	
11	912.4	82.6	1047	7	US-10-038-307-21	Sequence 21, Appl	
12	867.6	78.6	1823	7	US-10-038-307-11	Sequence 11, Appl	
13	853.4	77.3	5220	1	PCT-US02-08253-186	Sequence 186, App	
14	853.4	77.3	5220	1	PCT-US02-08253-300	Sequence 300, App	
15	787	71.3	1436	5	US-09-629-469A-11705	Sequence 11705, A	
16	521	47.2	1650	7	US-10-047-542-98	Sequence 98, Appl	
17	521	47.2	6802	7	US-10-047-542-100	Sequence 100, App	
18	453.2	41.1	4417	7	US-10-038-307-3	Sequence 3, Appl	
19	442.2	40.1	569	5	US-09-918-995-22508	Sequence 22508, A	
20	410	37.1	410	5	US-09-919-002-11267	Sequence 11267, A	
21	346.2	31.4	2234	7	US-10-104-047-669	Sequence 669, App	
22	344.6	31.2	3877	7	US-10-038-307-5	Sequence 5, Appl	
23	341.2	31.2	4089	8	US-60-373-595-9	Sequence 9, Appl	
24	341.2	30.9	2239	7	US-10-125-540-47	Sequence 47, Appl	
25	340.2	30.8	3501	7	US-10-038-307-7	Sequence 7, Appl	

```
QY 301 caaatccgtcaaggcctagaaagctccagaagtcttgcggggagagacacttacatg 360
Db 444 caaatccgtcaaggcctagaaagctccagaagtcttgcggggagagacacttacatg 503
QY 361 catgaaggatttgaaggccagtagcagatttatatgaacacagacaaaggttacag 420
Db 504 catgaaggatttgaaggccagtagcagatttatatgaacacagacaaaggttacag 563
QY 421 acagccagcgtcatcattcttctgactgatgagaaactccatgaagatctctttctat 480
Db 564 acagccagcgtcatcattcttctgactgatgagaaactccatgaagatctctttctat 623
QY 481 tcagagaggaggtcaataggctcagatcttgggcaattgtttactgtgtgtgtg 540
Db 624 tcagagaggaggtcaataggctcagatcttgggcaattgtttactgtgtgtgtg 683
QY 541 aaagatttcaatgagacacagctggcccgattgcggacagtagaagatcatgtgtttcc 600
Db 684 aaagatttcaatgagacacagctggcccgattgcggacagtagaagatcatgtgtttcc 743
QY 601 gtgaatgacggctttcaggctctgcaaggcatcatccactcaatttttgaagaagtcctgc 660
Db 744 gtgaatgacggctttcaggctctgcaaggcatcatccactcaatttttgaagaagtcctgc 803
QY 661 atcgaaattctagcagctgaacatccaccatattgtgcaggagatcatttcaagttgtc 720
Db 804 atcgaaattctagcagctgaacatccaccatattgtgcaggagatcatttcaagttgtc 863
QY 721 gtgagagaaacggcttccgacatgcccgaacgtgacaggggtcctctgcagcttcaag 780
Db 864 gtgagagaaacggcttccgacatgcccgaacgtgacaggggtcctctgcagcttcaag 923
QY 781 atcaatgactcgttcacactcaatgagaagcccttttctgtggaagatcttactgt 840
Db 924 atcaatgactcgttcacactcaatgagaagcccttttctgtggaagatcttactgt 983
QY 841 tctccagcgcctatcttataaagaagttggcatgaaagctgcactccaggtcagcatgaac 900
Db 984 tctccagcgcctatcttataaagaagttggcatgaaagctgcactccaggtcagcatgaac 1043
```

```
RESULT 2
PCT-US02-08253-231
; Sequence 231, Application PC/TUS0208253
; GENERAL INFORMATION:
; APPLICANT: Carson-Walter, Eleanor
; APPLICANT: St. Croix, Brad
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00179
; CURRENT APPLICATION NUMBER: PCT/US02/08253
; PRIOR APPLICATION NUMBER: 2002-04-10
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/308,829
```

```
; PRIOR FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 359
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 231
; LENGTH: 5540
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-08253-231

Query Match          98.8%; Score 1090.8; DB 1; Length 5540;
Best Local Similarity 99.8%; Pred. No. 1.2e-299;
Matches 1092; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atggccacggcgagcgagagagccctcgcatcgcgctccagtggtctctctttggccact 60
Db 144 atggccacggcgagcgagagagccctcgcatcgcgctccagtggtctctctttggccact 203
QY 61 ctgggtgctcatctgcgcgggcaagggggacgagggaggtgggtccagccttgctac 120
Db 204 ctgggtgctcatctgcgcgggcaagggggacgagggaggtgggtccagccttgctac 263
QY 121 ggcgatttgacctgtacttctatttggacaaatcagggaagtgtgtgcaccactggaaat 180
Db 264 ggcgatttgacctgtacttctatttggacaaatcagggaagtgtgtgcaccactggaaat 323
QY 181 gaaatctattactttgtggaacagttgggtccacaaatcattcagccacagttgagaatg 240
Db 324 gaaatctattactttgtggaacagttgggtccacaaatcattcagccacagttgagaatg 383
QY 241 tcttttattgttttctccacggaggaacaccttaataaactgacagaagacagagaa 300
Db 384 tcttttattgttttctccacggaggaacaccttaataaactgacagaagacagagaa 443
QY 301 caaatccgtcaaggcctagaaagctccagaagttctgcaggagagagacactttacatg 360
Db 444 caaatccgtcaaggcctagaaagctccagaagttctgcaggagagagacactttacatg 503
QY 361 catgaggagatttgaaggccagtagcagatttatatgaacacagacaaaggttacag 420
Db 504 catgaggagatttgaaggccagtagcagatttatatgaacacagacaaaggttacag 563
QY 421 acagccagcgtcatcattcttctgactgatgagaaactccatgaagatctctttctat 480
Db 564 acagccagcgtcatcattcttctgactgatgagaaactccatgaagatctctttctat 623
QY 481 tcagagaggaggtcaataggctcagatcttgggtgcaattgtttactgtgtgtgtg 540
Db 624 tcagagaggaggtcaataggctcagatcttgggtgcaattgtttactgtgtgtgtg 683
QY 541 aaagatttcaatgagacacagctggcccgattgcggacagtagaagatcatgtgtttccc 600
Db 684 aaagatttcaatgagacacagctggcccgattgcggacagtagaagatcatgtgtttccc 743
QY 601 gtgaatgacggctttcaggctctgcaaggcatccactcaatttttgaagaagtcctgc 660
Db 744 gtgaatgacggctttcaggctctgcaaggcatccactcaatttttgaagaagtcctgc 803
QY 661 atcgaaattctagcagctgaacatccaccatattgtgcaggagatcatttcaagttgtc 720
Db 804 atcgaaattctagcagctgaacatccaccatattgtgcaggagatcatttcaagttgtc 863
QY 721 gtgagagaaacggcttccgacatgcccgaacgtgacaggggtcctctgcagcttcaag 780
Db 864 gtgagagaaacggcttccgacatgcccgaacgtgacaggggtcctctgcagcttcaag 923
QY 781 atcaatgactcgttcacactcaatgagaagcccttttctgtggaagacacttattactg 840
Db 924 atcaatgactcgttcacactcaatgagaagcccttttctgtggaagacacttattactg 983
QY 841 tctccagcgcctatcttataaagaagttggcatgaaagctgcactccaggtcagcatgaac 900
Db 984 tctccagcgcctatcttataaagaagttggcatgaaagctgcactccaggtcagcatgaac 1043
```



Db	433	acagccagcgctcatctggtttgactgtagtgaggaaactccatgaaagatctcttttctctat	492
Qy	481	tcagagaggaggcctaataaggtctcgagatctctggtgcaattgtttactgtgtgtgtg	540
Db	493	tcagagaggaggcctaataaggtctcgagatctctggtgcaattgtttactgtgtgtgtg	552
Qy	541	aaagatttccaattgagacacacagctggcccggatctgggacagtgaagatcatgtgtttccc	600
Db	553	aaagatttccaattgagacacacagctggcccggatctgggacagtgaagatcatgtgtttccc	612
Qy	601	gtgaatgacggctttcaggctctgcaaggcatcatccaatcaattttgaaagaagtcctgc	660
Db	613	gtgaatgacggctttcaggctctgcaaggcatcatccaatcaattttgaaagaagtcctgc	672
Qy	661	atcgaattcttagcagctgaaccatccacatatgtgcaggagagtcatttcaagtgttc	720
Db	673	atcgaattcttagcagctgaaccatccacatatgtgcaggagagtcatttcaagtgttc	732
Qy	721	gtgagagaaacggcttccgcacatgcccgcaacgctggacaggctcctctgcagcttcaag	780
Db	733	gtgagagaaacggcttccgcacatgcccgcaacgctggacaggctcctctgcagcttcaag	792
Qy	781	atcaatgactcgtgcacactcaatgagaagcccttttctgtgagaagacactttattactg	840
Db	793	atcaatgactcgtgcacactcaatgagaagcccttttctgtgagaagacactttattactg	852
Qy	841	tgtccagcgccctatcttaaaagaagttggcatgaaagctgcactccaggctcagcatgaac	900
Db	853	tgtccagcgccctatcttaaaagaagttggcatgaaagctgcactccaggctcagcatgaac	912
Qy	901	gatggcctctcttttatctccagttctctgctcatcataccaccacacactgttctgacggt	960
Db	913	gatggcctctcttttatctccagttctctgctcatcataccaccacacactgttctgacggt	972
Qy	961	tcca 964	
Db	973	ccc 976	

```

RESULT      4
US-10-038-307-13
; Sequence 13, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZRAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-038-307-13

Query Match      86.3%; Score 952.4; DB 7; Length 1650;
Best Local Similarity 99.4%; Pred. No. 1.7e-260;
Matches 956; Conservative 0; Mismatches 6; Indels 0; Gaps

Qy      1  atggccacggcggagcggagagccctcgggcagcgcagcaggagatgggggtccagctgctac 120
Db      1  atggccacggcggagcggagagccctcgggcagcgcagcaggagatgggggtccagctgctac 120
Qy      61  ctgtgtctcactcgcggcggaaggagccctcgggcagcgcagcaggagatgggggtccagctgctac 120
Db      61  ctgtgtctcactcgcggcggaaggagccctcgggcagcgcagcaggagatgggggtccagctgctac 120
Qy      121  qgcgqatttgccctgtactctattttgacaaaatcaaaagatgtgctgcaccactggaat 120
Db      121  qgcgqatttgccctgtactctattttgacaaaatcaaaagatgtgctgcaccactggaat 120

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-038-307-15

Query Match      86.3%; Score 952.4; DB 7; Length 1650;
Best Local Similarity 99.4%; Pred. No. 1.7e-260;
Matches 956; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 atggccacggcgagcgagagccctcgccatcggtctccagtggtctctttggccaact 60
Db 1 atggccacggcgagcgagagccctcgccatcggtctccagtggtctctttggccaact 60
QY 61 ctgggtgctcatctgcgcgcaaggagcgagcgaggtggggtccagcctgtctac 120
Db 61 ctgggtgctcatctgcgcgcaaggagcgagcgaggtggggtccagcctgtctac 120
QY 121 ggcggatttgacctgtacttctatttggacaaatcagaaagtgtgctgcaccactggaaat 180
Db 121 ggcggatttgacctgtacttctatttggacaaatcagaaagtgtgctgcaccactggaaat 180
QY 181 gaaatctattacttgggaacagttggtcacaattcatcagccacagttgagaatg 240
Db 181 gaaatctattacttgggaacagttggtcacaattcatcagccacagttgagaatg 240
QY 241 tcctttattgttttcccccaggagaaacaccttaattgaaactgacagaagacagagaa 300
Db 241 tcctttattgttttcccccaggagaaacaccttaattgaaactgacagaagacagagaa 300
QY 301 caaatccgtcaaggcctagaaactccagaagaattctccagggagagacacttacatg 360
Db 301 caaatccgtcaaggcctagaaactccagaagaattctccagggagagacacttacatg 360
QY 361 catgaaggatttgaaggccagtgagcagatttattgaaacagacaagggtacagg 420
Db 361 catgaaggatttgaaggccagtgagcagatttattgaaacagacaagggtacagg 420
QY 421 acagccagcgctcatctgttcttgactgtggaactccatcgaagatctcttttctat 480
Db 421 acagccagcgctcatctgttcttgactgtggaactccatcgaagatctcttttctat 480
QY 481 tcagagagggaggtcaatagttctcagatcttggtgcaatttttactgtgtgtgtg 540
Db 481 tcagagagggaggtcaatagttctcagatcttggtgcaatttttactgtgtgtgtg 540
QY 541 aaagatttcaatgagacacagctggcccggtatgaggacagtaaggaatcatgtgttccc 600
Db 541 aaagatttcaatgagacacagctggcccggtatgaggacagtaaggaatcatgtgttccc 600
QY 601 gtgaatgacggcttccaggtctgcaaggcatccactcaattttgagaagatcctgc 660
Db 601 gtgaatgacggcttccaggtctgcaaggcatccactcaattttgagaagatcctgc 660
QY 661 atcgaaattctagcagctgaacctccaccataatgtaggagagatcatttcaagtgtc 720
Db 661 atcgaaattctagcagctgaacctccaccataatgtaggagagatcatttcaagtgtc 720
QY 721 gtgagagaaacggcttccgacatgcccgcaacgtgacagggccctctgcagcttcaag 780
Db 721 gtgagagaaacggcttccgacatgcccgcaacgtgacagggccctctgcagcttcaag 780
QY 781 atcaatgactcgttcacactcaatgagaagccctttctgtggaagacacttattactg 840
Db 781 atcaatgactcgttcacactcaatgagaagccctttctgtggaagacacttattactg 840
QY 841 tgtccagcgcttatctaaagaagttggcatgaaagctgcactccaggtcagatgaac 900
Db 841 tgtccagcgcttatctaaagaagttggcatgaaagctgcactccaggtcagatgaac 900
QY 901 gatggcctctctttatctccagttctgcatcatccaccacacacactgttctgcaggt 960
Db 901 gatggcctctctttatctccagttctgcatcatccaccacacacactgttctgcaggt 960

QY 961 tc 962
Db 961 tc 962

RESULT 5
US-10-038-307-15
; Sequence 15, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-038-307-15

Query Match      86.3%; Score 952.4; DB 7; Length 1650;
Best Local Similarity 99.4%; Pred. No. 1.7e-260;
Matches 956; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 atggccacggcgagcgagagccctcgccatcggtctccagtggtctctttggccaact 60
Db 1 atggccacggcgagcgagagccctcgccatcggtctccagtggtctctttggccaact 60
QY 61 ctgggtgctcatctgcgcgcaaggagcgagcgaggtggggtccagcctgtctac 120
Db 61 ctgggtgctcatctgcgcgcaaggagcgagcgaggtggggtccagcctgtctac 120
QY 121 ggcggatttgacctgtacttctatttggacaaatcagaaagtgtgctgcaccactggaaat 180
Db 121 ggcggatttgacctgtacttctatttggacaaatcagaaagtgtgctgcaccactggaaat 180
QY 181 gaaatctattacttgggaacagttggtcacaattcatcagccacagttgagaatg 240
Db 181 gaaatctattacttgggaacagttggtcacaattcatcagccacagttgagaatg 240
QY 241 tcctttattgttttcccccaggagaaacaccttaattgaaactgacagaagacagagaa 300
Db 241 tcctttattgttttcccccaggagaaacaccttaattgaaactgacagaagacagagaa 300
QY 301 caaatccgtcaaggcctagaaactccagaagaattctccagggagagacacttacatg 360
Db 301 caaatccgtcaaggcctagaaactccagaagaattctccagggagagacacttacatg 360
QY 361 catgaaggatttgaaggccagtgagcagatttattgaaacagacaagggtacagg 420
Db 361 catgaaggatttgaaggccagtgagcagatttattgaaacagacaagggtacagg 420
QY 421 acagccagcgctcatctgttcttgactgtggaactccatcgaagatctcttttctat 480
Db 421 acagccagcgctcatctgttcttgactgtggaactccatcgaagatctcttttctat 480
QY 481 tcagagagggaggtcaatagttctcagatcttggtgcaatttttactgtgtgtgtg 540
Db 481 tcagagagggaggtcaatagttctcagatcttggtgcaatttttactgtgtgtgtg 540
QY 541 aaagatttcaatgagacacagctggcccggtatgaggacagtaaggaatcatgtgttccc 600
Db 541 aaagatttcaatgagacacagctggcccggtatgaggacagtaaggaatcatgtgttccc 600
QY 601 gtgaatgacggcttccaggtctgcaaggcatccactcaattttgagaagatcctgc 660
Db 601 gtgaatgacggcttccaggtctgcaaggcatccactcaattttgagaagatcctgc 660
QY 661 atcgaaattctagcagctgaacctccaccataatgtaggagagatcatttcaagtgtc 720
Db 661 atcgaaattctagcagctgaacctccaccataatgtaggagagatcatttcaagtgtc 720
QY 721 gtgagagaaacggcttccgacatgcccgcaacgtgacagggccctctgcagcttcaag 780
Db 721 gtgagagaaacggcttccgacatgcccgcaacgtgacagggccctctgcagcttcaag 780
QY 781 atcaatgactcgttcacactcaatgagaagccctttctgtggaagacacttattactg 840
Db 781 atcaatgactcgttcacactcaatgagaagccctttctgtggaagacacttattactg 840
QY 841 tgtccagcgcttatctaaagaagttggcatgaaagctgcactccaggtcagatgaac 900
Db 841 tgtccagcgcttatctaaagaagttggcatgaaagctgcactccaggtcagatgaac 900
QY 901 gatggcctctctttatctccagttctgcatcatccaccacacacactgttctgcaggt 960
Db 901 gatggcctctctttatctccagttctgcatcatccaccacacacactgttctgcaggt 960
```

```
|||||
Db 901 gatggcctctttatctccagttctgtcatcatcaccacacacactgtttcccaaa 960
Qy 961 tc 962
||
Db 961 tc 962

RESULT 6
US-10-038-307-9
; Sequence 9, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 137 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-038-307-9
```

```
Query Match 86.1%; Score 950.8; DB 7; Length 1650;
Best Local Similarity 99.3%; Pred. No. 4.9e-260;
Matches 955; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 atggccacggcgagcgagagccctcgcatcggtccagttccagtggtctctttggccact 60
Db 1 atggccacggcgagcgagagccctcgcatcggtccagttccagtggtctctttggccact 60

Qy 61 ctgggtctcatctgcgcggcgaaggggacgcagggaggtgggggtccagcctgctac 120
Db 61 ctgggtctcatctgcgcggcgaaggggacgcagggaggtgggggtccagcctgctac 120

Qy 121 ggcggattgacctgtactctattttgacaaatcaggaagtgtgtgcacacactggaat 180
Db 121 ggcggattgacctgtactctattttgacaaatcaggaagtgtgtgcacacactggaat 180

Qy 181 gaaatctattcttgtgaacagttgctcacaattcatcagccacagattgagaatg 240
Db 181 gaaatctattcttgtggaacagttgctcacaattcatcagccacagattgagaatg 240

Qy 241 tcctttattgtttctccaccggagaaaccttaataagaactgacagaagacagaaa 300
Db 241 tcctttattgtttctccaccggagaaaccttaataagaactgacagaagacagaaa 300

Qy 301 caaatccgtcaaggcctgaagaacctccagaaagtctgccaggaggagacacttacatg 360
Db 301 caaatccgtcaaggcctgaagaacctccagaaagtctgccaggaggagacacttacatg 360

Qy 361 catgaagatttgaagggccagtgagcagatttattatgaacacagaagggtacagg 420
Db 361 catgaagatttgaagggccagtgagcagatttattatgaacacagaagggtacagg 420

Qy 421 acagccagctcatctgttctgactgatggagaactccatgaagatctcttttctat 480
Db 421 acagccagctcatctgttctgactgatggagaactccatgaagatctcttttctat 480

Qy 481 taagagagggaggcctaagttctcgagatcttgggtgcaattgtttactgtgtgtgtg 540
Db 481 taagagagggaggcctaagttctcgagatcttgggtgcaattgtttactgtgtgtgtg 540

Qy 541 aaagatttcaatgaacacagctggccggattgcccagtaagatcatgtgtttccc 600
Db 541 aaagatttcaatgaacacagctggccggattgcccagtaagatcatgtgtttccc 600
```

```
Qy 601 gtgaatgacgcttttcaggctctcaaggcatcaccactcaatttttgaagaagctctgc 660
Db 601 gtgaatgacgcttttcaggctctcaaggcatcaccactcaatttttgaagaagctctgc 660

Qy 661 atcgaaattctagcagctgaaccatccaccatagtgcaggagagtcatttcaagtgttc 720
Db 661 atcgaaattctagcagctgaaccatccaccatagtgcaggagagtcatttcaagtgttc 720

Qy 721 gtgagagaaacggctctccgacatgccgcaacatgaacagggctcctctcagctcaag 780
Db 721 gtgagagaaacggctctccgacatgccgcaacatgaacagggctcctctcagctcaag 780

Qy 781 atcaatgactcgtcacactcaatgagagccctttcttctgtggaagacacttattactg 840
Db 781 atcaatgactcgtcacactcaatgagagccctttcttctgtggaagatactattactg 840

Qy 841 tgtccagcgcctattttaaaagaagttgcatgaaagtgcactccaggtcagcatgaac 900
Db 841 tgtccagcgcctattttaaaagaagttgcatgaaagtgcactccaggtcagcatgaac 900

Qy 901 gatggcctcttttattctccagttctgtcatcatcaccacacacactgttctgacggt 960
Db 901 gatggcctcttttattctccagttctgtcatcatcaccacacacactgttcttccccaaa 960

Qy 961 tc 962
||
Db 961 tc 962
```

```
RESULT 7
US-10-038-307-23
; Sequence 23, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 1056
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-038-307-23

Query Match 86.0%; Score 949.4; DB 7; Length 1056;
Best Local Similarity 99.9%; Pred. No. 1e-259;
Matches 950; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atggccacggcgagcgagagccctcgcatcggtctccagtggtctcttttggccact 60
Db 13 atggccacggcgagcgagagccctcgcatcggtctccagtggtctcttttggccact 72

Qy 61 ctggtgtctatctgcgcggcgcaaggggacgcaggggagtggggtgccagcctgctac 120
Db 73 ctggtgtctatctgcgcggcgcaaggggacgcaggggagtggggtgccagcctgctac 132

Qy 121 ggcggattgacctgtactctattttgacaaatcaggaagtgtgtgcacacactggaat 180
Db 133 ggcggattgacctgtactctattttgacaaatcaggaagtgtgtgcacacactggaat 192

Qy 181 gaaatctattcttgtggaacagttgctcacaattcatcagccacagttgagaatg 240
Db 193 gaaatctattcttgtggaacagttgctcacaattcatcagccacagttgagaatg 252

Qy 241 tcctttattgtttctccaccggagaaaccttaataagaactgacagaagacagaaa 300
Db 253 tcctttattgtttctccaccggagaaaccttaataagaactgacagaagacagaaa 312
```

QY	301	caaatccgtcaaggccatagaagaactccagaagaagtctctccaggaggagagacacttacatg	360
Db	313	caaatccgtcaaggccatagaagaactccagaagaagtctctccaggaggagagacacttacatg	372
QY	361	catgaaggatttgaaggccagtgagcagatttattatgaataacagacaaaggtacagg	420
Db	373	catgaaggatttgaaggccagtgagcagatttattatgaataacagacaaaggtacagg	432
QY	421	acagccagcgtcatatgcttgaactgatgagaaactccatgaagatctcttttctat	480
Db	433	acagccagcgtcatatgcttgaactgatgagaaactccatgaagatctcttttctat	492
QY	481	tcagagaggagcgtcaatagctctcagatcttggtgcaattgttttaactgtgtgtgtg	540
Db	493	tcagagaggagcgtcaatagctctcagatcttggtgcaattgttttaactgtgtgtgtg	552
QY	541	aaagatttcaatgagacacagctggcccgattgagcagtaagtaagtaagtaagtaag	600
Db	553	aaagatttcaatgagacacagctggcccgattgagcagtaagtaagtaagtaagtaag	612
QY	601	gtgaatgacggcttccagctcgcagcagcagcagcagcagcagcagcagcagcagcag	660
Db	613	gtgaatgacggcttccagctcgcagcagcagcagcagcagcagcagcagcagcagcag	672
QY	721	gtgagagaaacggcttccagctcgcagcagcagcagcagcagcagcagcagcagcagcag	780
Db	733	gtgagagaaacggcttccagctcgcagcagcagcagcagcagcagcagcagcagcagcag	792
QY	841	tgccagcgcctatcttaaaagaagttggatgaagcagcagcagcagcagcagcagcagcag	900
Db	853	tgccagcgcctatcttaaaagaagttggatgaagcagcagcagcagcagcagcagcagcag	912
QY	901	gatggcctcttttatctccagttctgtcatcatcaccacacacactgt	951
Db	913	gatggcctcttttatctccagttctgtcatcatcaccacacacactgt	963
RESULT			
US-10-038-307-1			
; Sequence 1, Application US/10038307			
; GENERAL INFORMATION:			
; APPLICANT: James B. ROTTMAN			
; APPLICANT: Theresa L. O'KEEFE			
; APPLICANT: Engin OZKAYNAK			
; APPLICANT: Judith J. HEALEY			
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods			
; FILE REFERENCE: 7853-253-999			
; CURRENT APPLICATION NUMBER: US/10/038,307			
; CURRENT FILING DATE: 2002-06-28			
; NUMBER OF SEQ ID NOS: 26			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 19			
; LENGTH: 1713			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-10-038-307-19			
Query Match 86.0%; Score 949.4; DB 7; Length 1713;			
Best Local Similarity 99.9%; Pred. No. 1.3e-259;			
Matches 950; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
QY	1	atggccacggcgagcgagagcctcgccagtcggtccagtggtctcttttggccact	60

```
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2272
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (213)...(1211)
US-10-038-307-1

Query Match      86.0%; Score 949.4; DB 7; Length 2272;
Best Local Similarity 99.9%; Pred. No. 1.4e-259;
Matches 950; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atggccacggcgagcgagagccctcgccatcgctccagtgctctctttggccact 60
Db 213 atggccacggcgagcgagagccctcgccatcgctccagtgctctctttggccact 272

Qy 61 ctggtgctcatctgcccgcggaaggacgacggagggatgggggtccagcctgctac 120
Db 273 ctggtgctcatctgcccgcggaaggacgacggagggatgggggtccagcctgctac 332

Qy 121 ggcggaattgacctgacttattttgacaaatcagaaagtgtgctgcaccactggaat 180
Db 333 ggcggaattgacctgacttattttgacaaatcagaaagtgtgctgcaccactggaat 392

Qy 181 gaaatctattactttgtggaacagttgctcacaaatccatcagccacagttgagaatg 240
Db 393 gaaatctattactttgtggaacagttgctcacaaatccatcagccacagttgagaatg 452

Qy 241 tctttattgttttccaccgcgaggaacaaccttaatgaaactgacagaagacagagaa 300
Db 453 tctttattgttttccaccgcgaggaacaaccttaatgaaactgacagaagacagagaa 512

Qy 301 caaatccgctcaaggcctagaagaactccagaaagtctgccagagagagacacttacatg 360
Db 513 caaatccgctcaaggcctagaagaactccagaaagtctgccagagagagacacttacatg 572

Qy 361 catgaagatttgaaaggccagtgagcagattttattgaaaaacagacaagggtcacagg 420
Db 573 catgaagatttgaaaggccagtgagcagattttattgaaaaacagacaagggtcacagg 632

Qy 421 acagccagcgctcatctgcttggactgatggagaactccatgaagatctcttttctat 480
Db 633 acagccagcgctcatctgcttggactgatggagaactccatgaagatctcttttctat 692

Qy 481 tcagagagggaggtctaataaggctctcgagatcttggcgaattgttactgtgtgtgtg 540
Db 693 tcagagagggaggtctaataaggctctcgagatcttggcgaattgttactgtgtgtgtg 752

Qy 541 aaagatttcaattgagacacagctggcccgattgcgacagtaaggaatcatgtgtttccc 600
Db 753 aaagatttcaattgagacacagctggcccgattgcgacagtaaggaatcatgtgtttccc 812

Qy 601 gtgaatgacggctttcaggctctgaaggcatcaccactcaattttgaagaagtcctgc 660
Db 813 gtgaatgacggctttcaggctctgaaggcatcaccactcaattttgaagaagtcctgc 872

Qy 661 atcgaattctagcagctgaaccatccaccatagtgcaggagagctcattcaagtgtgc 720
Db 873 atcgaattctagcagctgaaccatccaccatagtgcaggagagctcattcaagtgtgc 932

Qy 721 gtgagagaaacggcttcgcagatgcccgcaacgctggacagggctcctcgagcttcaag 780
Db 933 gtgagagaaacggcttcgcagatgcccgcaacgctggacagggctcctcgagcttcaag 992

Qy 781 atcaatgactcgttcacactcaatgagaagccctttctgtgtggaagacacttattactg 840
Db 993 atcaatgactcgttcacactcaatgagaagccctttctgtgtggaagacacttattactg 1052
```

```
Qy 841 tgtccagcgccctatctttaaaagaattggcatgaaagctgcactccaggtcagcatgaac 900
Db 1053 tgtccagcgccctatctttaaaagaattggcatgaaagctgcactccaggtcagcatgaac 1112

Qy 901 gatggcctctcttttatctccagttctgtcatcatcaccacacacactgtg 951
Db 1113 gatggcctctcttttatctccagttctgtcatcatcaccacacacactgtg 1163

RESULT 10
US-10-038-307-25
; Sequence 25, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 1008
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-038-307-25
```

```
Query Match      86.0%; Score 949; DB 7; Length 1008;
Best Local Similarity 99.5%; Pred. No. 1.3e-259;
Matches 952; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 atggccacggcgagcgagagccctcgccatcgctccagtgctctctttggccact 60
Db 13 atggccacggcgagcgagagccctcgccatcgctccagtgctctctttggccact 72

Qy 61 ctggtgctcatctgcccgcggaaggacgacggagggatgggggtccagcctgctac 120
Db 73 ctggtgctcatctgcccgcggaaggacgacggagggatgggggtccagcctgctac 132

Qy 121 ggcggaattgacctgacttattttgacaaatcagaaagtgtgctgcaccactggaat 180
Db 133 ggcggaattgacctgacttattttgacaaatcagaaagtgtgctgcaccactggaat 192

Qy 181 gaaatctattactttgtggaacagttgctcacaaatccatcagccacagttgagaatg 240
Db 193 gaaatctattactttgtggaacagttgctcacaaatccatcagccacagttgagaatg 252

Qy 241 tctttattgttttccaccgcgaggaacaaccttaatgaaactgacagaagacagagaa 300
Db 253 tctttattgttttccaccgcgaggaacaaccttaatgaaactgacagaagacagagaa 312

Qy 301 caaatccgctcaaggcctagaagaactccagaaagtctgccagggagagacacttacatg 360
Db 313 caaatccgctcaaggcctagaagaactccagaaagtctgccagggagagacacttacatg 372

Qy 361 catgaagatttgaaaggccagtgagcagattttattgaaaaacagacaagggtcacagg 420
Db 373 catgaagatttgaaaggccagtgagcagattttattgaaaaacagacaagggtcacagg 432

Qy 421 acagccagcgctcatctgcttggactgatggagaactccatgaagatctcttttctat 480
Db 433 acagccagcgctcatctgcttggactgatggagaactccatgaagatctcttttctat 492

Qy 481 tcagagagggaggtctaataaggctctcgagatcttggcgaattgttactgtgtgtgtg 540
Db 493 tcagagagggaggtctaataaggctctcgagatcttggcgaattgttactgtgtgtgtg 552

Qy 541 aaagatttcaattgagacacagctggcccgatttggcgacagtaaggaatcatgtgtttccc 600
Db 541 aaagatttcaattgagacacagctggcccgatttggcgacagtaaggaatcatgtgtttccc 600
```



```

QY 65 tgctcatctgcgcgcgaaggagagcagcagggaggggtccagcgtctacggcg 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38 taccacaattgtccgtccggggacgcagggaggtccagcgtctacggcg 97
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 125 gatttgacctgtacttcatcttttgacaaatcagaaagtgtcgtcaccactggaatgaaa 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 98 gatttgacctgtacttcatcttttgacaaatcagaaagtgtcgtcaccactggaatgaaa 157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 185 tctattacttcttggaacagttggtccacaaattctacgccccacagttgagaatgtcct 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 158 tctattacttcttggaacagttggtccacaaattctacgccccacagttgagaatgtcct 217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 245 ttattgtttctccaccgcggaggaacaccttaattgaactgacagagacagagacaaa 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 218 ttattgtttctccaccgcggaggaacaccttaattgaactgacagagacagagacaaa 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 305 tccgtcaaggccttagaagaactccagaagaatttctgcaggaggagacacttacatgcatg 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 278 tccgtcaaggccttagaagaactccagaagaatttctgcaggaggagacacttacatgcatg 337
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 365 aaggatttgaaagggccagtgagcagatttatatgaaacagacagaaggtacaggacag 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 338 aaggatttgaaagggccagtgagcagatttatatgaaacagacagaaggtacaggacag 397
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 425 ccagcgtcatcattgtttgactgatggagaactccatgaagatctcttttctattcag 484
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 398 ccagcgtcatcattgtttgactgatggagaactccatgaagatctcttttctattcag 457
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 485 agaggaggctaataagttctcgagatcttggtgcaattgtttactgtgttggtgtaag 544
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 458 agaggaggctaataagttctcgagatcttggtgcaattgtttactgtgttggtgtaag 517
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 545 atttcaatgagacacagctgcccggattgcgacagtaaggatcatgtttccctgta 604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 518 atttcaatgagacacagctgcccggattgcgacagtaaggatcatgtgtttccctgta 577
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 605 atcagcgtcttcaggctctcaaggcatcatccactcaattttgaaagtcctgcacg 664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 578 atcagcgtcttcaggctctcaaggcatcatccactcaattttgaaagtcctgcacg 637
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 665 aaattctagcagctgaacatccaccatagtgcaggagagtgatcttcaagtgtcgtga 724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 638 aaattctagcagctgaacatccaccatagtgcaggagagtgatcttcaagtgtcgtga 697
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 725 gaggaaacggtctccgacatgccgcgaactggacaggttccctgcagcttcaagatca 784
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 698 gaggaaacggtctccgacatgccgcgaactggacaggttccctgcagcttcaagatca 757
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 785 atgactcgttcacactcaatgagaagccctttctgtggaagacacttattactgtgtc 844
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 758 atgactcgttcacactcaatgagaagccctttctgtggaagacacttattactgtgtc 817
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 845 cagcgcctattctaaagaagttggaatgaaagctgcactccaggtcagatgaacgatg 904
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 818 cagcgcctattctaaagaagttggaatgaaagctgcactccaggtcagatgaacgatg 877
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 905 gctctcttttattctcagttctgtctcatcatcaccacacacactgtctgacggttc 962
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 878 gctctcttttattctcagttctgtctcatcatcaccacacacactgtcttctccaaatc 935
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 13

```

PCT-US02-08253-186
: Sequence 186, Application PC/TUS0208253
: GENERAL INFORMATION:
: APPLICANT: Carson-Walter, Eleanor
: APPLICANT: St. Croix, Brad
: APPLICANT: Vogelstein, Bert
: APPLICANT: Kinzler, Kenneth
: TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
: FILE REFERENCE: 1107 00179
: CURRENT APPLICATION NUMBER: PCT/US02/08253
: CURRENT FILING DATE: 2002-04-10

```

```

: PRIOR APPLICATION NUMBER: 60/282,850
: PRIOR FILING DATE: 2001-04-11
: PRIOR APPLICATION NUMBER: 60/308,829
: PRIOR FILING DATE: 2001-08-01
: NUMBER OF SEQ ID NOS: 359
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 186
: LENGTH: 5220
: TYPE: DNA
: ORGANISM: Mus musculus
PCT-US02-08253-186

```

```

Query Match 77.3%; Score 853.4; DB 1; Length 5220;
Best Local Similarity 87.3%; Pred. No. 3.8e-232;
Matches 935; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 24 cctcgccatcggttccagctgtctcttttggccactctggtctcatctgcgcggcga 83
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291 cctgggtgcgggctgcgggactctgctggctgcactgtgctgtgtgcgcggaca 350
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 84 agggggacgcaggagagatgggggtccagcctgctacggcgatttgacctgtactcat 143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 351 cgggggcccgcggcaggtgggggaccagcttgcacggaggattccacctactcat 410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 144 ttgtgacaaatcagaaagtgtgctgcaccactggaaatgaaatctattacttttggaa 203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 411 cctggacaaatcagaaagtgtgctgcaccactggaaatgaaatctactacttgcggaga 470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 204 gttggtccacaaattcatcagccacagtcacagtgatgtccctttattgttttctccacc 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 471 gttggtccatagattcatcagccacagtcacagtgatgtccctttattgttttctactcg 530
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 264 aggaacaaccttaataaactgacagaagacagagacaaatccgtcaagcctagaaga 323
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 531 agggacaaacttaataaactaactgaggacagggagacagatccgcacaaagcctcagaaga 590
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 324 actccagaaagtctccaggagagacacttcatgataaggatttgaaagggccag 383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 591 gctccagaaagtctccaggagagacacttcatgataaggatttcgagagggccag 650
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 384 tgagcagatttattgaaacacagaaagggtacagacagccagcgctcatctgtctt 443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 651 tgagcagatttattgaaacacagaaagggtacagacagcgagcgctcatctgcgtt 710
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 444 gactgatggaaactccatgaagatctcttttctatttattcagagagggaggaataggtc 503
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 711 gacggatggggagctgcacgagggacctcttcttactcagagggagggtcaccgcatc 770
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 504 tcgagactctgtgcaattgtttactgtgtgtgtgaaagatttcaatgagacacagct 563
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 771 ccgagacctgtgtcgattgtttactgcgttggtgcgtgaaggatttcaatgaaactcagtt 830
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 564 ggcccggattgcgcagcagtaaggatcatgtgtttcccggtgaatgacgggttccaggtct 623
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 831 ggctcggattgcagacagtaaggaccgtgtttctgtgtaacgacggttccaggtct 890
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 624 gcaaggcatcatccactcaattttgaaagactcctgcatacgaataatctcagcagctgaacc 683
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 891 ccaaggcattatccactcaatttttaagaaatcctgcatacgaataatctcggcggtgaacc 950
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 684 atccaccatatgtgcaggagagtgatcttcaagtgtgtgtgagaggaacacggttcccgaca 743
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 951 atccaccatatgtgcaggagagtgatcttcaagtgtgtgtgagaggaacacggttcccgaca 1010
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 744 tgcccgcacagtggaacgggttcctctgcagcttcaagatcaaatgactcgggtcacaactcaa 803
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1011 tgcccgcacagtggaacgggttcctctgcagcttcaagatcaaatgactcagtcacgctcaa 1070
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 804 tgagaagccctttctgtggaagacacttattactgtgtccagcgcctatcttaagaaga 863
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1071 tgagaagccctttctgtggaagacacttattactgtgtccagcgcctatcttaagaaga 1130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```





; PRIOR APPLICATION NUMBER: JP 2000-241899
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/159,590
; PRIOR FILING DATE: 1999-10-18
; PRIOR APPLICATION NUMBER: 60/183,322
; PRIOR FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 19025
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11705
; LENGTH: 1436
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (380)..(1033)
US-09-629-469A-11705

Query Match 71.3%; Score 787; DB 5; Length 1436;
Best Local Similarity 99.3%; Pred. No. 1.8e-213;
Matches 801; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Qy 1 atggccacgcgagcgagagccctcgcatcggtctccagtggtctctttggccact 60
Db 144 atggccacgcgagcgagagccctcgcatcggtctccagtggtctct-acggtccact 202
Qy 61 ctggtgctcatctgcgcgggcaaggggagcgagggatgggggtccagcctgctac 120
Db 203 ctggtgctcatctgcgcgggcaaggggagcgagggatgggggtccagcctgctac 262
Qy 121 ggcggatttgacctgtacttctattttggacaaatcaggaaagtgtgctgcaccactggaaat 180
Db 263 ggcggatttgacctgtacttctattttggacaaatcaggaaagtgtgctgcaccactggaaat 322
Qy 181 gaaatctattactttgtggaaagttgggttcacaaattcatcagccacagttgagaatg 240
Db 323 gaaatctattactttgtggaaagttgggttcacaaattcatcagccacagttgagaatg 382
Qy 241 tcccttattgtttctccaccggaggaacaccttaataaactgacagagacagagaa 300
Db 383 tcccttattgtttctccaccggaggaacaccttaataaactgacagagacagagaa 442
Qy 301 caaatccgtcaagccttagaagaactccagaaagtctgccaggagggagacacttacatg 360
Db 443 caaatccgtcaagccttagaagaactccagaaagtctgccaggagggagacacttacatg 502
Qy 361 catgaaggatttgaaagggccagtgagcagatttattatgaaacacagacaggggtacagg 420
Db 503 catgaaggatttgaaagggccagtgagcagatttattatgaaacacagacaggggtacagg 562
Qy 421 acagccagcgtcatctgttgaactgtgagaaactccatgaagatctcttttctat 480
Db 563 acagctagcgtcatctgttgaactgtgagaaactccatgaagatctcttttctat 622
Qy 481 tcagagagggaggttaataagttctcgagatcttgggtcaattgtttactgtgtgtgtg 540
Db 623 tcagagagggaggttaataagttctcgagatcttgggtcaattgtttactgtgtgtgtg 682
Qy 541 aaagatttcaatgagacacagctggccggattgcggacagtaaggatcatgtgttccc 600
Db 683 aaagatttcaatgagacacagctggccggattgcggacagtaaggatcatgtgttccc 742
Qy 601 gtgaatcacggtttcaggtctcgaaggcatcatcactcaattttgaaagacctgc 660
Db 743 gtgaatcacggtttcaggtctcgaaggcatcatcactcaattttgaaagacctgc 802
Qy 661 atcgaattcttagcagctgaaccatccaccatattgtcaggagagcatcttcaagtgtc 720
Db 803 atcgaattcttagcagctgaaccatccaccatattgtcaggagagcatcttcaagtgtc 862
Qy 721 gtgagaggaacacggttccgacatgcccgcaacgtggacaggggtcctctgcagcttcaag 780
Db 863 gtgagaggaacacggttccgacatgcccgcaacgtggacaggggtcctctgcagcttcaag 922

Qy 781 atcaatgactcggtcacactcaatgag 807
Db 923 atcaatgactcggtcacactcagtaag 949
Search completed: August 9, 2002, 02:03:04
Job time: 10801 sec

us-09-970-076-1\_copy\_104\_1207.rnpn

Fri Aug 9 10:56:58 2002

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 9, 2002, 01:53:53 ; Search time 4997.37 Seconds  
(without alignments)

4779.500 Million cell updates/sec

Title: us-09-970-076-1\_copy\_104\_1207

Perfect score:

Sequence: 1104

1 atggccacgagcgagcgag.....gtgaggaaataataaaaa 1104

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 21979536 seqs, 10817449327 residues

Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_NA\_Main :

1: /cgn2\_6/ptodata/2/pna/PCRTUS\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/pna/US06\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/pna/US07\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/pna/US080\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/pna/US081\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/pna/US082\_COMB.seq.\*  
7: /cgn2\_6/ptodata/2/pna/US083\_COMB.seq.\*  
8: /cgn2\_6/ptodata/2/pna/US084\_COMB.seq.\*  
9: /cgn2\_6/ptodata/2/pna/US085\_COMB.seq.\*  
10: /cgn2\_6/ptodata/2/pna/US086\_COMB.seq.\*  
11: /cgn2\_6/ptodata/2/pna/US087\_COMB.seq.\*  
12: /cgn2\_6/ptodata/2/pna/US088\_COMB.seq.\*  
13: /cgn2\_6/ptodata/2/pna/US089\_COMB.seq.\*  
14: /cgn2\_6/ptodata/2/pna/US090\_COMB.seq.\*  
15: /cgn2\_6/ptodata/2/pna/US091\_COMB.seq.\*  
16: /cgn2\_6/ptodata/2/pna/US092\_COMB.seq.\*  
17: /cgn2\_6/ptodata/2/pna/US093\_COMB.seq.\*  
18: /cgn2\_6/ptodata/2/pna/US094\_COMB.seq.\*  
19: /cgn2\_6/ptodata/2/pna/US095A\_COMB.seq.\*  
20: /cgn2\_6/ptodata/2/pna/US095B\_COMB.seq.\*  
21: /cgn2\_6/ptodata/2/pna/US095C\_COMB.seq.\*  
22: /cgn2\_6/ptodata/2/pna/US095D\_COMB.seq.\*  
23: /cgn2\_6/ptodata/2/pna/US095A\_COMB.seq.\*  
24: /cgn2\_6/ptodata/2/pna/US095B\_COMB.seq.\*  
25: /cgn2\_6/ptodata/2/pna/US095C\_COMB.seq.\*  
26: /cgn2\_6/ptodata/2/pna/US095D\_COMB.seq.\*  
27: /cgn2\_6/ptodata/2/pna/US095E\_COMB.seq.\*  
28: /cgn2\_6/ptodata/2/pna/US095F\_COMB.seq.\*  
29: /cgn2\_6/ptodata/2/pna/US095G\_COMB.seq.\*  
30: /cgn2\_6/ptodata/2/pna/US095H\_COMB.seq.\*  
31: /cgn2\_6/ptodata/2/pna/US095I\_COMB.seq.\*  
32: /cgn2\_6/ptodata/2/pna/US095J\_COMB.seq.\*  
33: /cgn2\_6/ptodata/2/pna/US095K\_COMB.seq.\*  
34: /cgn2\_6/ptodata/2/pna/US095L\_COMB.seq.\*  
35: /cgn2\_6/ptodata/2/pna/US095M\_COMB.seq.\*  
36: /cgn2\_6/ptodata/2/pna/US095N\_COMB.seq.\*  
37: /cgn2\_6/ptodata/2/pna/US095O\_COMB.seq.\*  
38: /cgn2\_6/ptodata/2/pna/US095P\_COMB.seq.\*  
39: /cgn2\_6/ptodata/2/pna/US095Q\_COMB.seq.\*  
40: /cgn2\_6/ptodata/2/pna/US095R\_COMB.seq.\*  
41: /cgn2\_6/ptodata/2/pna/US095S\_COMB.seq.\*  
42: /cgn2\_6/ptodata/2/pna/US095T\_COMB.seq.\*  
43: /cgn2\_6/ptodata/2/pna/US095U\_COMB.seq.\*

44: /cgn2\_6/ptodata/2/pna/US6005\_COMB.seq.\*  
45: /cgn2\_6/ptodata/2/pna/US6006\_COMB.seq.\*  
46: /cgn2\_6/ptodata/2/pna/US6007\_COMB.seq.\*  
47: /cgn2\_6/ptodata/2/pna/US6008\_COMB.seq.\*  
48: /cgn2\_6/ptodata/2/pna/US6009\_COMB.seq.\*  
49: /cgn2\_6/ptodata/2/pna/US6010\_COMB.seq.\*  
50: /cgn2\_6/ptodata/2/pna/US6011\_COMB.seq.\*  
51: /cgn2\_6/ptodata/2/pna/US6012\_COMB.seq.\*  
52: /cgn2\_6/ptodata/2/pna/US6013\_COMB.seq.\*  
53: /cgn2\_6/ptodata/2/pna/US6014\_COMB.seq.\*  
54: /cgn2\_6/ptodata/2/pna/US6015\_COMB.seq.\*  
55: /cgn2\_6/ptodata/2/pna/US6016\_COMB.seq.\*  
56: /cgn2\_6/ptodata/2/pna/US6017\_COMB.seq.\*  
57: /cgn2\_6/ptodata/2/pna/US6018\_COMB.seq.\*  
58: /cgn2\_6/ptodata/2/pna/US6019\_COMB.seq.\*  
59: /cgn2\_6/ptodata/2/pna/US6020\_COMB.seq.\*  
60: /cgn2\_6/ptodata/2/pna/US6021\_COMB.seq.\*  
61: /cgn2\_6/ptodata/2/pna/US6022\_COMB.seq.\*  
62: /cgn2\_6/ptodata/2/pna/US6023\_COMB.seq.\*  
63: /cgn2\_6/ptodata/2/pna/US6024\_COMB.seq.\*  
64: /cgn2\_6/ptodata/2/pna/US6025\_COMB.seq.\*  
65: /cgn2\_6/ptodata/2/pna/US6026\_COMB.seq.\*  
66: /cgn2\_6/ptodata/2/pna/US6027\_COMB.seq.\*  
67: /cgn2\_6/ptodata/2/pna/US6028\_COMB.seq.\*  
68: /cgn2\_6/ptodata/2/pna/US6029\_COMB.seq.\*  
69: /cgn2\_6/ptodata/2/pna/US6030\_COMB.seq.\*  
70: /cgn2\_6/ptodata/2/pna/US6031\_COMB.seq.\*  
71: /cgn2\_6/ptodata/2/pna/US6032\_COMB.seq.\*  
72: /cgn2\_6/ptodata/2/pna/US6033\_COMB.seq.\*  
73: /cgn2\_6/ptodata/2/pna/US6034\_COMB.seq.\*  
74: /cgn2\_6/ptodata/2/pna/US6035\_COMB.seq.\*  
75: /cgn2\_6/ptodata/2/pna/US6036\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1104	100.0	1414	36	US-09-970-076-1	Sequence 1, Appl
2	1090.8	98.8	2386	60	US-60-213-359-1222	Sequence 1222, Ap
3	1090.8	98.8	5540	34	US-09-918-715-176	Sequence 176, App
4	1090.8	98.8	5540	34	US-09-918-715-231	Sequence 231, App
5	1090.8	98.8	5540	36	US-09-970-076-5	Sequence 5, Appl
6	1090.6	98.8	2447	1	PCT-US00-30045-14	Sequence 14, Appl
7	1089.4	98.7	2086	1	PCT-US00-30045-45	Sequence 45, Appl
8	949.4	86.0	999	1	PCT-US99-31025-10	Sequence 10, Appl
9	949.4	86.0	999	16	US-09-223-546-10	Sequence 10, Appl
10	949.4	86.0	999	18	US-09-471-179-10	Sequence 10, Appl
11	949.4	86.0	999	22	US-09-599-596-10	Sequence 10, Appl
12	949.4	86.0	2112	36	US-09-970-076-7	Sequence 7, Appl
13	949.4	86.0	2272	1	PCT-US99-31025-8	Sequence 8, Appl
14	949.4	86.0	2272	16	US-09-223-546-8	Sequence 8, Appl
15	949.4	86.0	2272	18	US-09-471-179-8	Sequence 8, Appl
16	949.4	86.0	2272	22	US-09-599-596-8	Sequence 8, Appl
17	949.4	86.0	2272	30	US-09-796-753-11	Sequence 11, Appl
18	949.4	86.0	2374	25	US-09-644-871-9264	Sequence 9264, Ap
19	949.4	86.0	2374	28	US-09-710-281-4357	Sequence 4357, Ap
20	949.4	86.0	2459	17	US-09-329-548A-68	Sequence 68, Appl
21	949.4	86.0	2459	18	US-09-419-553-204	Sequence 204, App
22	949.4	86.0	2459	19	US-09-516-448-998	Sequence 998, App
23	949.4	86.0	2459	30	US-09-770-155-204	Sequence 204, App
24	949.4	86.0	2459	30	US-09-780-183-68	Sequence 68, Appl
25	949.4	86.0	2459	32	US-09-833-381-998	Sequence 998, App
26	949.4	86.0	2459	32	US-09-833-382-998	Sequence 998, App
27	947.8	85.9	2401	17	US-09-371-168-7727	Sequence 7727, Ap
28	947.8	85.9	2401	25	US-09-644-868-7637	Sequence 7637, Ap
29	947.8	85.9	2401	25	US-09-644-871-7309	Sequence 7309, Ap
30	947.8	85.9	2401	25	US-09-649-163-6281	Sequence 6281, Ap
31	947.8	85.9	2401	25	US-09-652-121-5593	Sequence 5593, Ap

```
32 947.8 85.9 2401 25 US-09-652-127-7873 Sequence 7873, Ap
33 947.8 85.9 2401 31 US-09-801-833-7727 Sequence 7727, Ap
34 937.4 84.9 2485 26 US-09-668-337-188 Sequence 188, Ap
35 860 77.9 2397 24 US-09-625-102-1757 Sequence 1757, Ap
36 853.4 77.3 5220 34 US-09-918-715-186 Sequence 186, Ap
37 853.4 77.3 5220 34 US-09-918-715-300 Sequence 300, Ap
38 829.8 75.2 2981 75 US-60-360-207-20845 Sequence 20845, A
39 803.8 72.8 1402 18 US-09-471-275-9977 Sequence 9977, Ap
40 803.8 72.8 1402 18 US-09-488-725A-3907 Sequence 3907, Ap
41 803.8 72.8 1609 24 US-09-488-725A-335 Sequence 335, Ap
42 803.8 72.8 1609 24 US-09-620-312B-8 Sequence 8, Appli
43 803.8 72.8 1609 37 US-10-037-270-8 Sequence 8, Appli
44 787 71.3 1436 36 US-09-970-076-9 Sequence 9, Appli
45 516.8 46.8 3981 30 US-09-799-451-250 Sequence 250, App
```

ALIGNMENTS

```
RESULT 1
US-09-970-076-1
; Sequence 1, Application US/09970076
; GENERAL INFORMATION:
; APPLICANT: Young, John A.T.
; APPLICANT: Bradley, Kenneth A.
; APPLICANT: Collier, Robert J.
; APPLICANT: Mogridge, Jeremy S.
; TITLE OF INVENTION: Anthrax Toxin Receptor
; FILE REFERENCE: 960296.97745
; CURRENT APPLICATION NUMBER: US/09/970,076
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: 60/251,481
; PRIOR FILING DATE: 2000-12-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1414
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (104)..(1207)
US-09-970-076-1
```

```
Query Match 100.0%; Score 1104; DB 36; Length 1414;
Best Local Similarity 100.0%; Pred. No. 2.4e-281;
Matches 1104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgggccacggcgagcgagagccctcgccatcggtccagtcggctctcttggccact 60
Dbb 104 atgggccacggcgagcgagagccctcgccatcggtccagtcggctctcttggccact 163

QY 61 ctggtgtcatctgcccgggcaaggggacgcagggagatgggggtccagcctgctac 120
Dbb 164 ctggtgtcatctgcccgggcaaggggacgcagggagatgggggtccagcctgctac 223

QY 121 ggcggattgacctgactctatttggacaaatcaggaagtgtctgcaccactggaat 180
Dbb 224 ggcggattgacctgactctatttggacaaatcaggaagtgtctgcaccactggaat 283

QY 181 gaaatctattcttggaaacagttgggttcacaaattcattcagccacagttgagaatg 240
Dbb 284 gaaatctattcttggaaacagttgggttcacaaattcattcagccacagttgagaatg 343

QY 241 tcctttatgttttccaccggagacaaaccttaataaactgacagaagacagagaa 300
Dbb 344 tcctttatgttttccaccggagacaaaccttaataaactgacagaagacagagaa 403

QY 301 caaatccgtaaggcctagagaactccagaaagtctcgaggaggagacacttacatg 360
Dbb 404 caaatccgtaaggcctagagaactccagaaagtctcgaggaggagacacttacatg 463
```

```
QY 361 catgaagatttgaaggccagtgagcagatttatgaacaacagacaaggtacagg 420
Dbb 464 catgaagatttgaaggccagtgagcagatttatgaacaacagacaaggtacagg 523
QY 421 acagccagcgtcatcattgtttgactgatggagaacctcattgaagatctctttctat 480
Dbb 524 acagccagcgtcatcattgtttgactgatggagaacctcattgaagatctctttctat 583
QY 481 tcagagaggagcgttaattaggtctcgagatcttggtgcaattgtttactgtgttggtg 540
Dbb 584 tcagagaggagcgttaattaggtctcgagatcttggtgcaattgtttactgtgttggtg 643
QY 541 aagaatttcattgagacacagctggccggattgctggacagtaaggatcattgtttccc 600
Dbb 644 aagaatttcattgagacacagctggccggattgctggacagtaaggatcattgtttccc 703
QY 601 gtgaatgacggcttccaggtctctgcaaggcatccacatccactcaatttgaagaagctgc 660
Dbb 704 gtgaatgacggcttccaggtctctgcaaggcatccacatccactcaatttgaagaagctgc 763
QY 661 atcgaaattctagcagctgaaccatccaccatattgtgcaggagagtcatttcaagtgtc 720
Dbb 764 atcgaaattctagcagctgaaccatccaccatattgtgcaggagagtcatttcaagtgtc 823
QY 721 gtgagagaaacggcttccagatgccgcgaacgtggacagggctcctctgcagcttcaag 780
Dbb 824 gtgagagaaacggcttccagatgccgcgaacgtggacagggctcctctgcagcttcaag 883
QY 781 atcaatgactcgttcacactcaatgaagagccctttctgtggaagacacttattactg 840
Dbb 884 atcaatgactcgttcacactcaatgaagagccctttctgtggaagacacttattactg 943
QY 841 tgtccagcgcctattctaaagaagttggcatgaaagtgcactccaggtgcagcatgaac 900
Dbb 944 tgtccagcgcctattctaaagaagttggcatgaaagtgcactccaggtgcagcatgaac 1003
QY 901 gatggcctctctttatctccagttctgtatcaccacacacacactgtctgacggt 960
Dbb 1004 gatggcctctctttatctccagttctgtatcaccacacacacactgtctgacggt 1063
QY 961 tccatcctggccatcgccctgctgacctgtctctgctcctagccctggctctctctgg 1020
Dbb 1064 tccatcctggccatcgccctgctgacctgtctctgctcctagccctggctctctctgg 1123
QY 1021 tggttctggccctctgctgactgtgattatcaaggaggtccctccacccctgcagag 1080
Dbb 1124 tggttctggccctctgctgactgtgattatcaaggaggtccctccacccctgcagag 1183
QY 1081 gagagtgagaaaaataaaaaa 1104
Dbb 1184 gagagtgagaaaaataaaaaa 1207
```

```
RESULT 2
US-60-213-359-1222
; Sequence 1222, Application US/60213359
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Poly
; TITLE OF INVENTION: Identified Thereby
; FILE REFERENCE: GX-0015 P
; CURRENT APPLICATION NUMBER: US/60/213,359
; CURRENT FILING DATE: 2000-06-21
; NUMBER OF SEQ ID NOS: 7924
; SOFTWARE: PERL Program
; SEQ ID NO 1222
; LENGTH: 2386
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No: 243308.2  
US-60-213-359-1222

Query Match 98.8%; Score 1090.8; DB 60; Length 2386;  
Best Local Similarity 99.8%; Pred. No. 1.2e-287;  
Matches 1092; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 atgcccacgagcgagcgagccctcgccatcgctccagctgctccagctgctctcttggccact 60  
Db 295 atgcccacgagcgagcgagccctcgccatcgctccagctgctccagctgctctcttggccact 354

Qy 61 ctggtgctcatctgcgcgggcaaggggagcgagggaggggtccagctgctctcttggccact 120  
Db 355 ctggtgctcatctgcgcgggcaaggggagcgagggaggggtccagctgctctcttggccact 414

Qy 121 ggcggattgacctgactctattttgacaaaatcaggaagtgtgctgcacactggaaat 180  
Db 415 ggcggattgacctgactctattttgacaaaatcaggaagtgtgctgcacactggaaat 474

Qy 181 gaactctattactgtggaacagtgtgctcacaataatccacacacacacagttgaagaatg 240  
Db 475 gaactctattactgtggaacagtgtgctcacaataatccacacacacacagttgaagaatg 534

Qy 241 tcccttattgtttctccaccggaggaacaaccttaataagaaactgacagaagacagagaa 300  
Db 535 tcccttattgtttctccaccggaggaacaaccttaataagaaactgacagaagacagagaa 594

Qy 301 caaatccgtcaagcgctagaaagactccagaaaagtctgccagaggagacacttacatg 360  
Db 595 caaatccgtcaagcgctagaaagactccagaaaagtctgccagaggagacacttacatg 654

Qy 361 catcaagattgaaagggcgagcgagcagatttatgaaacagaaaggggtacagg 420  
Db 655 catcaagattgaaagggcgagcgagcagatttatgaaacagaaaggggtacagg 714

Qy 421 acagccagcgtcatctgcttgaactgatggagaactccatgaagatctctttcttat 480  
Db 715 acagccagcgtcatctgcttgaactgatggagaactccatgaagatctctttcttat 774

Qy 481 tcagagaggggggtaataagctctcgagatcttggtgcaattgtttactgtgtgtgtg 540  
Db 775 tcagagaggggggtaataagctctcgagatcttggtgcaattgtttactgtgtgtgtg 834

Qy 541 aaagatttcaatgagacacagctgcccggagctgcggacagtaagatcatgtgtttccc 600  
Db 835 aaagatttcaatgagacacagctgcccggagctgcggacagtaagatcatgtgtttccc 894

Qy 601 gtgaatgacgctttcaggcctcgaagcctcaatccactcaattttgaaagaagtcctgc 660  
Db 895 gtgaatgacgctttcaggcctcgaagcctcaatccactcaattttgaaagaagtcctgc 954

Qy 661 atcgaattcttagcagctgaaccatccaccatattgtgcagagagctatttcaagtgtc 720  
Db 955 atcgaattcttagcagctgaaccatccaccatattgtgcagagagctatttcaagtgtc 1014

Qy 721 gtgagagaaacggcttccgacatgcccgcaacgtggagagggctcctctgcagcttcaag 780  
Db 1015 gtgagagaaacggcttccgacatgcccgcaacgtggagagggctcctctgcagcttcaag 1074

Qy 781 atcaatgactcgggtcacactcaatgagaagcccttttctgtggaagacactatttactg 840  
Db 1075 atcaatgactcgggtcacactcaatgagaagcccttttctgtggaagacactatttactg 1134

Qy 841 tgtccagcgctatatttaaaagaagtgtggcatgaaagtgcactccaggtcagcatgaac 900  
Db 1135 tgtccagcgctatatttaaaagaagtgtggcatgaaagtgcactccaggtcagcatgaac 1194

Qy 901 gatggcctctctttatctccagttctgtcatcatccacacacacactgttctgacggt 960  
Db 1195 gatggcctctctttatctccagttctgtcatcatccacacacacactgttctgacggt 1254

Qy 961 tccatcctggccatcgccctgctgatcctgttctctgctcctagcctgctctctctctgg 1020  
Db 1255 tccatcctggccatcgccctgctgatcctgttctctgctcctagcctgctctctctctgg 1314

Qy 1021 tggttctggccccctctgctgactgtgattatcaaggagggtccctccacccccctgcccag 1080  
Db 1315 tggttctggccccctctgctgactgtgattatcaaggagggtccctccacccccctgcccag 1374

Qy 1081 gagagtggagaaaa 1094  
Db 1375 gagagtggagaaaa 1388

RESULT 3  
US-09-918-715-176  
; Sequence 176 Application US/09918715  
; GENERAL INFORMATION:  
; APPLICANT: Brad St. Croix  
; APPLICANT: Bert Vogelstein  
; APPLICANT: Kenneth Kinzler  
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
; FILE REFERENCE: 1107.00134  
; CURRENT APPLICATION NUMBER: US/09/918,715  
; PRIOR FILING DATE: 2001-08-01  
; PRIOR APPLICATION NUMBER: 60/222,599  
; PRIOR FILING DATE: 2000-08-02  
; PRIOR APPLICATION NUMBER: 60/224,360  
; PRIOR FILING DATE: 2000-08-11  
; PRIOR APPLICATION NUMBER: 60/282,850  
; PRIOR FILING DATE: 2000-04-11  
; NUMBER OF SEQ ID NOS: 358  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 176  
; LENGTH: 5540  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-918-715-176

Query Match 98.8%; Score 1090.8; DB 34; Length 5540;  
Best Local Similarity 99.8%; Pred. No. 1.7e-287;  
Matches 1092; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 atggccacgagcgagcgagagccctcgccatcgctccagcttccagtgctctctttggccact 60  
Db 144 atggccacgagcgagcgagagccctcgccatcgctccagtgctctctttggccact 203

Qy 61 ctggtgctcatctgcgcgggcaaggggagcgagggaggggtccagctgctctctctctctac 120  
Db 204 ctggtgctcatctgcgcgggcaaggggagcgagggaggggtccagctgctctctctac 263

Qy 121 ggcggattgacctgactctattttgacaaaatcaggaagtgtgctgcacactggaaat 180  
Db 264 ggcggattgacctgactctattttgacaaaatcaggaagtgtgctgcacactggaaat 323

Qy 181 gaaatctattactttgtggaacagtgtgctcacaataatccacacacacagttggaagt 240  
Db 324 gaaatctattactttgtggaacagtgtgctcacaataatccacacacagttggaagt 383

Qy 241 tcccttattgtttctccaccggaggaacaaccttaataagaaactgacagaagacagagaa 300  
Db 384 tcccttattgtttctccaccggaggaacaaccttaataagaaactgacagaagacagagaa 443

Qy 301 caaatccgtcaagcgcttagaagaactccagaaaagtcttgcagaggagagacacttacatg 360  
Db 444 caaa ccccgcaagggcctgagaagaactccagaaaagtcttgcagaggagacacttacatg 503

Qy 361 catgaagatttgaagggccagtgagcagattttattatgaaaaacagacaaggggtacagg 420  
Db 504 catgaagatttgaagggccagtgagcagattttattatgaaaaacagacaaggggtacagg 563

Qy 421 acagccagcgtcatctgcttggactgatggagaactccatgagaagatctctttcttat 480  
Db 421 acagccagcgtcatctgcttggactgatggagaactccatgagaagatctctttcttat 480

Fri Aug 9 10:56:57 2002

Db 564 acagccagcgtcatcattgtcttgactgatggagaacctccatgaagatctcttttttat 623  
Qy 481 tcagagagagagcgaataatggtctcagatcttggcgaattgtttactgtgtgtgtg 540  
Db 624 tcagagagagagcgaataatggtctcagatcttggcgaattgtttactgtgtgtgtg 683  
Qy 541 aaagattcaatgagacacagctggcccgattggcagacagtaagatcatgtgtttccc 600  
Db 684 aaagattcaatgagacacagctggcccgattggcagacagtaagatcatgtgtttccc 743  
Qy 601 gtgaatgaagcgtttcaggcttcgaagcagcatcatccatcaatttgaagaagtctgc 660  
Db 744 gtgaatgaagcgtttcaggcttcgaagcagcatcatccatcaatttgaagaagtctgc 803  
Qy 661 atcgaaattctagcagctgaacatccaccatagtgcaggagagtcatttcaagtgtc 720  
Db 804 atcgaaattctagcagctgaacatccaccatagtgcaggagagtcatttcaagtgtc 863  
Qy 721 gtgagagaaacggtctccagatcccgcaactggacaggtctctctgcagcttcaag 780  
Db 864 gtgagagaaacggtctccagatcccgcaactggacaggtctctctgcagcttcaag 923  
Qy 781 atcaatgactcgttcacactcaatgagaagcccttttctgtggaagacacttattactg 840  
Db 924 atcaatgactcgttcacactcaatgagaagcccttttctgtggaagacacttattactg 983  
Qy 841 tgtccagcgcctactctaaagaagttggcatgaaagtcaactcaggtcagcatgaac 900  
Db 984 tgtccagcgcctactctaaagaagttggcatgaaagtcaactcaggtcagcatgaac 1043  
Qy 901 gatggcctcttttatctcagttctgtcatcatcatcaccacacacactgttctgacggt 960  
Db 1044 gatggcctcttttatctcagttctgtcatcatcatcaccacacacactgttctgacggt 1103  
Qy 961 tccatctctggcactcgcctgtgatcctgtctgtctgtctgtctgtctgtctgtctgtg 1020  
Db 1104 tccatctctggcactcgcctgtgatcctgtctgtctgtctgtctgtctgtctgtctgtg 1163  
Qy 1021 tggttctggccctctgcctgcactgtattatcaagaggtccctccaccctgccgag 1080  
Db 1164 tggttctggccctctgcctgcactgtattatcaagaggtccctccaccctgccgag 1223  
Qy 1081 gagagtgaggaaaa 1094  
Db 1224 gagagtgaggaaaa 1237

RESULT 4  
US-09-918-715-231  
; Sequence 231, Application US/09918715  
; GENERAL INFORMATION:  
; APPLICANT: Brad St. Croix  
; APPLICANT: Bert Vogelstein  
; APPLICANT: Kenneth Kinzler  
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
; FILE REFERENCE: 1107.00134  
; CURRENT APPLICATION NUMBER: 2001-08-01  
; CURRENT FILING DATE: 2001-08-01  
; PRIOR APPLICATION NUMBER: 60/222,599  
; PRIOR FILING DATE: 2000-08-02  
; PRIOR APPLICATION NUMBER: 60/224,360  
; PRIOR FILING DATE: 2000-08-11  
; PRIOR APPLICATION NUMBER: 60/282,850  
; PRIOR FILING DATE: 2000-04-11  
; NUMBER OF SEQ ID NOS: 358  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 231  
; LENGTH: 5540  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-918-715-231

Query Match 98.8%; Score 1090.8; DB 34; Length 5540;  
Best Local Similarity 99.8%; Pred. No. 1.7e-287;  
Matches 1092; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 atggccacggcggagcggagagccctcgccatcgcttccagtggtctcttttggccact 60  
Db 144 atggccacggcggagcggagcggcctcgccatcgcttccagtggtctcttttggccact 203  
Qy 61 ctggtgctcatctgcgcgggcaaggggacgagggaggatgggggtccagcctgtctac 120  
Db 204 ctggtgctcatctgcgcgggcaaggggacgagggaggatgggggtccagcctgtctac 263  
Qy 121 ggcgatttgacctgtacttacttatttgacaaatcaggaaagtgtctgcacacactggaat 180  
Db 264 ggcgatttgacctgtacttacttatttgacaaatcaggaaagtgtctgcacacactggaat 323  
Qy 181 gaaatctattacttctgtggaacagttgtgtcacaatctcatcagccacagttgagaatg 240  
Db 324 gaaatctattacttctgtggaacagttgtgtcacaatctcatcagccacagttgagaatg 383  
Qy 241 tctttattgttttctccaccggaggaaacacttaattgaaactgacagaacacagagaa 300  
Db 384 tctttattgttttctccaccggaggaaacacttaattgaaactgacagaacacagagaa 443  
Qy 301 caaatccgtaaggcctagaaacactccagaaagtctctccaggaggagacacacttacatg 360  
Db 444 caaatccgtaaggcctagaaacactccagaaagtctctccaggaggagacacacttacatg 503  
Qy 361 catgaaggatttgaaggccaggtgagcagatttattatgaaaacacagaaaggtacag 420  
Db 504 catgaaggatttgaaggccaggtgagcagatttattatgaaaacacagaaaggtacag 563  
Qy 421 acagcagcgtctcatcattgtttgactgatgagaactcatcagaatctctttttctat 480  
Db 564 acagcagcgtctcatcattgtttgactgatgagaactcatcagaatctctttttctat 623  
Qy 481 tcagagagggagcgttaataagttctcgagatcttggtgcaattgtttactgtgtgtgtg 540  
Db 624 tcagagagggagcgttaataagttctcgagatcttggtgcaattgtttactgtgtgtgtg 683  
Qy 541 aaagatttcaatgagacacacagctggcccggtatgcgacagtaagatcatgtgttccc 600  
Db 684 aaagatttcaatgagacacacagctggcccggtatgcgacagtaagatcatgtgttccc 743  
Qy 601 gtgaatgacggtttcaggctctgcaaggcatcatccactcaattttgagaagaagtcctgc 660  
Db 744 gtgaatgacggtttcaggctctgcaaggcatcatccactcaattttgagaagaagtcctgc 803  
Qy 661 atcgaaattcttagcagctgaacatccaccatagtgcaggagagtcatttcaagtgtc 720  
Db 804 atcgaaattcttagcagctgaacatccaccatagtgcaggagagtcatttcaagtgtc 863  
Qy 721 gtgagagaaacggtctccgacatgccgcaactggacaggtctctctgcagcttcaag 780  
Db 864 gtgagagaaacggtctccgacatgccgcaactggacaggtctctctgcagcttcaag 923  
Qy 781 atcaatgactcgttcacactcaatgagaagcccttttctgtggaagacacttattactg 840  
Db 924 atcaatgactcgttcacactcaatgagaagcccttttctgtggaagacacttattactg 983  
Qy 841 tgtccagcgcctacttataaagaagttggcatgaaagtgcactccaggtcagcatgaac 900  
Db 984 tgtccagcgcctacttataaagaagttggcatgaaagtgcactccaggtcagcatgaac 1043  
Qy 901 gatggcctcttttatctcagttctgtcatcatcaccacacacactgttctgacggt 960  
Db 1044 gatggcctcttttatctcagttctgtcatcatcaccacacacactgttctgacggt 1103  
Qy 961 tccatctctggcactcgcctgtgatcctgtctgtctgtctgtctgtctgtctgtctgtg 1020  
Db 1104 tccatctctggcactcgcctgtgatcctgtctgtctgtctgtctgtctgtctgtctgtg 1163  
Qy 1021 tggttctggccctctgcctgcactgtattatcaagaggttcctccaccctgccgag 1080

Db 1164 tgggttcgccccctctgtcgaactgtgattatcaaggaggtccctccaccctcgcgag 1223  
 Qy 1081 gagagtgaggaaaa 1094  
 Db 1224 gagagtgaggaaaa 1237

RESULT 5  
 US-09-970-076-5  
 ; Sequence 5, Application US/09970076  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Young, John A.T.  
 ; APPLICANT: Bradley, Kenneth A.  
 ; APPLICANT: Collier, Robert J.  
 ; APPLICANT: Mogridge, Jeremy S.  
 ; TITLE OF INVENTION: Anthrax Toxin Receptor  
 ; FILE REFERENCE: 960296.97745  
 ; CURRENT APPLICATION NUMBER: US/09/970, 076  
 ; CURRENT FILING DATE: 2001-10-03  
 ; PRIOR APPLICATION NUMBER: 60/251,481  
 ; PRIOR FILING DATE: 2000-12-05  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 5  
 ; LENGTH: 5540  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (144)..(1835)  
 US-09-970-076-5

Query Match 98.8%; Score 1090.8; DB 36; Length 5540;  
 Best Local Similarity 99.8%; Pred. No. 1.7e-287;  
 Matches 1092; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 atggccacggcgagcgagagccctcgcatcggtctccagtggtctctttggccact 60  
 Db 144 atggccacggcgagcgagagccctcgcatcggtctccagtggtctctttggccact 203  
 Qy 61 ctggtgctcatctgcgcggcggaaggcgaggggagggagggaggggtccagcctgtac 120  
 Db 204 ctggtgctcatctgcgcggcggaaggcgaggggagggaggggtccagcctgtac 263  
 Qy 121 ggcggattgaactgtacttcttcaatttggacaaatcaggaaagtgtcgtcaccactggaat 180  
 Db 264 ggcggattgaactgtacttcttcaatttggacaaatcaggaaagtgtcgtcaccactggaat 323  
 Qy 181 gaaatctattactttgtggaacagttggtctcaaaaattcatcagccacacagttgagaatg 240  
 Db 324 gaaatctattactttgtggaacagttggtctcaaaaattcatcagccacacagttgagaatg 383  
 Qy 241 tccattattgttttctccaccggagggaacacaccttaataatgaaactgacagaagacagagaa 300  
 Db 384 tccattattgttttctccaccggagggaacacaccttaataatgaaactgacagaagacagagaa 443  
 Qy 301 caaatccgctcaagccctagagaactccagaagaagtctgcagggagagacacttacatg 360  
 Db 444 caaatccgctcaagccctagagaactccagaagaagtctgcagggagagacacttacatg 503  
 Qy 361 catgaagatttgaaggccagtgagacagatttatgaaaaacagacaaagggtacagg 420  
 Db 504 catgaagatttgaaggccagtgagacagatttatgaaaaacagacaaagggtacagg 563  
 Qy 421 acagccagcgtcatctattgctttgactgatggagaactccatgaagatctctttttctat 480  
 Db 564 acagccagcgtcatctattgctttgactgatggagaactccatgaagatctctttttctat 623  
 Qy 481 tcagagagggagggctaataaggtctcagagatcttggtaattgttactgtgtgtgtgtg 540  
 Db 624 tcagagagggagggctaataaggtctcagagatcttggtaattgttactgtgtgtgtgtg 683

Qy 541 aaagatttcaatgagacacagctggcccgattgcgacagtaaggatcatgtgtttccc 600  
 Db 684 aaagatttcaatgagacacagctggcccgattgcgacagtaaggatcatgtgtttccc 743  
 Qy 601 gtgaatgacggctttcaggctctgcaaggcatcatccactcaattttgaagaagctctgc 660  
 Db 744 gtgaatgacggctttcaggctctgcaaggcatcatccactcaattttgaagaagctctgc 803  
 Qy 661 atcgaaattctagcagctgaaccatccaccatatgtgcaggagagtgatttcaagtgttc 720  
 Db 804 atcgaaattctagcagctgaaccatccaccatatgtgcaggagagtgatttcaagtgttc 863  
 Qy 721 gtgagagaaaacggcttccgacatgcccgaacgtggacagggctcctctgcagcttcaag 780  
 Db 864 gtgagagaaaacggcttccgacatgcccgaacgtggacagggctcctctgcagcttcaag 923  
 Qy 781 atcaatgactcgtgtcacactcaatgagaagcccttttctgtggaagacacttattactg 840  
 Db 924 atcaatgactcgtgtcacactcaatgagaagcccttttctgtggaagacacttattactg 983  
 Qy 841 tgtccagcgcctattctaaagaagtgtggcatgaaagtgcactccaggtcagcatgaac 900  
 Db 984 tgtccagcgcctattctaaagaagtgtggcatgaaagtgcactccaggtcagcatgaac 1043  
 Qy 901 gatggcctctcttttctccagttctgtcatcatcaccaccacacacactgttctgacgggt 960  
 Db 1044 gatggcctctcttttctccagttctgtcatcatcaccaccacacacactgttctgacgggt 1103  
 Qy 961 tccatcctggccatgcgccctgctgactgttctgtctcttagccctggctctctctctg 1020  
 Db 1104 tccatcctggccatgcgccctgctgactgttctgtctcttagccctggctctctctctg 1163  
 Qy 1021 tgggttggccctctctgctgcactgtgattatcaaggaggtccctccaccctgcgcgag 1080  
 Db 1164 tgggttggccctctctgctgcactgtgattatcaaggaggtccctccaccctgcgcgag 1223  
 Qy 1081 gagagtgaggaaaa 1094  
 Db 1224 gagagtgaggaaaa 1237

RESULT 6  
 PCT-US00-30045-14  
 ; Sequence 14, Application PC/TUS0030045  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Human Genome Sciences, Inc.  
 ; TITLE OF INVENTION: 28 Human Secreted Proteins  
 ; FILE REFERENCE: PS708PCT  
 ; CURRENT APPLICATION NUMBER: PCT/US00/30045  
 ; CURRENT FILING DATE: 2000-11-01  
 ; PRIOR APPLICATION NUMBER: 60/163,581  
 ; PRIOR FILING DATE: 1999-11-05  
 ; PRIOR APPLICATION NUMBER: 60/215,133  
 ; PRIOR FILING DATE: 2000-06-30  
 ; NUMBER OF SEQ ID NOS: 201  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 14  
 ; LENGTH: 2447  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 PCT-US00-30045-14

Query Match 98.8%; Score 1090.6; DB 1; Length 2447;  
 Best Local Similarity 99.6%; Pred. No. 1.4e-287;  
 Matches 1093; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 atggccacggcgagcgagagccctcgcatcggtctccagtggtctctttttggccact 60  
 Db 140 atggccacggcgagcgagagccctcgcatcggtctccagtggtctctttttggccact 199  
 Qy 61 ctggtgctcatctgcgcggcggaaggcgaggggagggaggggtccagcctgctac 120

Db 200 ctggttctcatctgcgcggaaggggacgagggagatgggggtccagcctgctac 259  
QY 121 ggcggatttgacctgtactctatttggacaatacaggagtgctgcaccactggaa 180  
Db 260 ggcggatttgacctgtactctatttggacaatacaggagtgctgcaccactggaa 319  
QY 181 gaaatattactttgtgaacagtggtgcacaaattcatcagccacagttgagaatg 240  
Db 320 gaaatattactttgtgaacagtggtgcacaaattcatcagccacagttgagaatg 379  
QY 241 tctttatttttctccacccaggaaacaccttaataaactgacagaagacagagaa 300  
Db 380 tctttatttttctccacccaggaaacaccttaataaactgacagaagacagagaa 439  
QY 301 caaatccgtcaagcctagaaactccagaagttctgcaggagagacacttacatg 360  
Db 440 caaatccgtcaagcctagaaactccagaagttctgcaggagagacacttacatg 499  
QY 361 catgaagatttgaaggccagtgagcagatttattatgaatacagacagaggtacag 420  
Db 500 catgaagatttgaaggccagtgagcagatttattatgaatacagacagaggtacag 559  
QY 421 acagccagcgtcatcattgttgcattgagtgagaaactccatgaagatctctttctat 480  
Db 560 acagccagcgtcatcattgttgcattgagtgagaaactccatgaagatctctttctat 619  
QY 481 tcagagaggagggttaataagttctcgagatcttggtgcaattgtttactgtgtgtg 540  
Db 620 tcagagaggagggttaataagttctcgagatcttggtgcaattgtttactgtgtgtg 679  
QY 541 aaagatttcaatgagacacagctggccgggattgcgacagtagagcatgtgttccc 600  
Db 680 aaagatttcaatgagacacagctggccgggattgcgacagtagagcatgtgttccc 739  
QY 601 gtgaatgacggttccaggtctgcagacatcccgcaacgtggagaggttctctgaggtc 660  
Db 740 gtgaatgacggttccaggtctgcagacatcccgcaacgtggagaggttctctgaggtc 799  
QY 661 atcgaattctagcagctgaacattccaccatattgagagagagagagagagagag 720  
Db 800 atcgaattctagcagctgaacattccaccatattgagagagagagagagagagag 859  
QY 721 gtgagagaaacggttccagacatcccgcaacgtggagaggttctctgaggtc 780  
Db 860 gtgagagaaacggttccagacatcccgcaacgtggagaggttctctgaggtc 919  
QY 781 atcaatgactcgttcacactcaatgagaagccctttctgtggaagacacttattactg 840  
Db 920 atcaatgactcgttcacactcaatgagaagccctttctgtggaagacacttattactg 979  
QY 841 tgtccagcgcctattctaaagaagtggcatgaaagctgcactccaggtcagcatgaac 900  
Db 980 tgtccagcgcctattctaaagaagtggcatgaaagctgcactccaggtcagcatgaac 1039  
QY 901 gatggcctcttttctccaggttctgtcatcatcaccacacacactgttctgacggt 960  
Db 1040 gatggcctcttttctccaggttctgtcatcatcaccacacacactgttctgacggt 1099  
QY 961 tccatcctggcactgcgcctgtgactgttctgtctctgactgacgtgctctctg 1020  
Db 1100 tccatcctggcactgcgcctgtgactgttctgtctctgactgacgtgctctctg 1159  
QY 1021 tggcttgcgcctctgtgcactgtgattatacgaagggttccctccacccctgcgag 1080  
Db 1160 tggcttgcgcctctgtgcactgtgattatacgaagggttccctccacccctgcgag 1219  
QY 1081 gagagtggagaaataa 1097  
Db 1220 gagagtggagtaagta 1236

PCT-US00-30045-45  
; Sequence 45, Application PC/TUS0030045  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: 28 Human Secreted Proteins  
; FILE REFERENCE: PS708PCT  
; CURRENT APPLICATION NUMBER: PCT/US00/30045  
; CURRENT FILING DATE: 2000-11-01  
; PRIOR APPLICATION NUMBER: 60/163,581  
; PRIOR FILING DATE: 1999-11-05  
; PRIOR APPLICATION NUMBER: 60/215,133  
; PRIOR FILING DATE: 2000-06-30  
; NUMBER OF SEQ ID NOS: 201  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 45  
; LENGTH: 2086  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (10)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (2070)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (2075)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (2079)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (2083)  
; OTHER INFORMATION: n equals a,t,g, or c  
; PCT-US00-30045-45

Query Match 98.7%; Score 1089.4; DB 1; Length 2086;  
Best Local Similarity 99.4%; Pred. No. 2.8e-287;  
Matches 1090; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 1 atggccacgagcggagagccctcgcatcggttccagtggtctcttggccact 60  
Db 166 atggccacgagcggagagccctcgcatcggttccagtggtctcttggccact 225  
QY 61 ctggtgctcatctgcgcgggcaaggagcagcagagagtggttccagcctgtctac 120  
Db 226 ctggtgctcatctgcgcgggcaaggagcagcagagagtggttccagcctgtctac 285  
QY 121 ggcggatttgacctgtacttctatttggacaatacaggaagtgtgtgacacactggaat 180  
Db 286 ggcggatttgacctgtacttctatttggacaatacaggaagtgtgtgacacactggaat 345  
QY 181 gaaatattactttgtggaaacagttgtgcacaaattcatcagccacagttgagaatg 240  
Db 346 gaaatattactttgtggaaacagttgtgcacaaattcatcagccacagttgagaatg 405  
QY 241 tccatttattttctccacccaggaaacaccttaataaactgacagaagacagagaa 300  
Db 406 tccatttattttctccacccaggaaacaccttaataaactgacagaagacagagaa 465  
QY 301 caaatccgtcaagcctagaaactccagaagttctgcaggagagacacttacatg 360  
Db 466 caaatccgtcaagcctagaaactccagaagttctgcaggagagacacttacatg 525  
QY 361 catgaagatttgaaggccagtgagcagatttattatgaatacagaggttacag 420  
Db 526 catgaagatttgaaggccagtgagcagatttattatgaatacagaggttacag 585  
QY 421 acagccagcgtcatcattgttgcattgagtaggaaactccatgaagatctctttctat 480  
Db 586 acagccagcgtcatcattgttgcattgagtaggaaactccatgaagatctctttctat 645



QY 481 tcagagaggaggaataaggtctcagagatcttggtgcaattgtttactgtgtgtgtg 540  
 Db 646 tcagagaggaggaataaggtctcagagatcttggtgcaattgtttactgtgtgtgtg 705  
 QY 541 aaagatttcaatgagacacagctggcccgaggttgagacagtaagagatcatgtgtttccc 600  
 Db 706 aaagatttcaatgagacacagctggcccgaggttgagacagtaagagatcatgtgtttccc 765  
 QY 601 gtgaatgacgcttcaggctctgcaaggcatcatccacccaatttgaagaactctgc 660  
 Db 766 gtgaatgacgcttcaggctctgcaaggcatcatccacccaatttgaagaactctgc 825  
 QY 661 atcgaaattcttagcagctgaaccatccaccatatgtgcagagagatcatttcaagtgtc 720  
 Db 826 atcgaaattcttagcagctgaaccatccaccatatgtgcagagagatcatttcaagtgtc 885  
 QY 721 gtgagagaaacggtctccgacatgcccgaacgtgacaggggtccctgcagcttcaag 780  
 Db 886 gtgagagaaacggtctccgacatgcccgaacgtgacaggggtccctgcagcttcaag 945  
 QY 781 atcaatgactcgtgcacactcaatgagaagcccttttctgtggaagacacttattactg 840  
 Db 946 atcaatgactcgtgcacactcaatgagaagcccttttctgtggaagacacttattactg 1005  
 QY 841 tgcacagcgcctactttaaagaagttggcatgaaagctgcactccaggtcagcatgaac 900  
 Db 1006 tgcacagcgcctactttaaagaagttggcatgaaagctgcactccaggtcagcatgaac 1065  
 QY 901 gatggcctcttttatctccagttctgtcatcatccaccacacactgttctgaaggt 960  
 Db 1066 gatggcctcttttatctccagttctgtcatcatccaccacacactgttctgaaggt 1125  
 QY 961 tccatctgcccctgcccctgctgactgttctctgctcagcctgagcctctctctg 1020  
 Db 1126 tccatctgcccctgcccctgctgactgttctctgctcagcctgagcctctctctg 1185  
 QY 1021 tgggttggcccctgctgactgtgattatcaaggaggtccctccaccctgcccag 1080  
 Db 1186 tgggttggcccctgctgactgtgattatcaaggaggtccctccaccctgcccag 1245  
 QY 1081 gagagtgaggaataa 1097  
 Db 1246 gagagtgaggaataa 1262

RESULT 8  
 PCT-US99-31025-10  
 ; Sequence 10, Application PC/TUS9931025  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Millennium Pharmaceuticals, Inc.  
 ; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS  
 ; TITLE OF INVENTION: ENCODING THEM  
 ; FILE REFERENCE: 7853-173-228  
 ; CURRENT APPLICATION NUMBER: PCT/US99/31025  
 ; EARLIER FILING DATE: 1999-12-23  
 ; EARLIER APPLICATION NUMBER: 09/223,546  
 ; EARLIER FILING DATE: 1998-12-30  
 ; NUMBER OF SEQ ID NOS: 135  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 10  
 ; LENGTH: 999  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 PCT-US99-31025-10

Query Match 86.0%; Score 949.4; DB 1; Length 999;  
 Best Local Similarity 99.9%; Pred. No. 5e-249;  
 Matches 950; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atggccacggcgagagagccctcggaatcggtccagtggtctcttcttggccact 60  
 Db 1 atggccacggcgagagagccctcggaatcggtccagtggtctcttcttggccact 60

QY 61 ctggtgtctatctgcgcggcaaggggagcgagggagatgggggtccagcctgctac 120  
 Db 61 ctggtgtctatctgcgcggcaaggggagcgagggagatgggggtccagcctgctac 120  
 QY 121 ggcggatttgacctgtacttctatttggacaaatcaggaagtgtgctgcaccactggaat 180  
 Db 121 ggcggatttgacctgtacttctatttggacaaatcaggaagtgtgctgcaccactggaat 180  
 QY 181 gaaatctattacttgtggaacagttggctcacaattcatcagccccacagttagaag 240  
 Db 181 gaaatctattacttgtggaacagttggctcacaattcatcagccccacagttagaag 240  
 QY 241 tccatttattgtttctccaccggaggaacaaccttaataaactgacagaagacagagaa 300  
 Db 241 tccatttattgtttctccaccggaggaacaaccttaataaactgacagaagacagagaa 300  
 QY 301 caaatccgtcaagccttagaagaactccagaaagtcttgcaggagagacacttaccatg 360  
 Db 301 caaatccgtcaagccttagaagaactccagaaagtcttgcaggagagacacttaccatg 360  
 QY 361 catgaagatttgaagggccagtgagcagatttattatgaaaaacagacaggggtacagg 420  
 Db 361 catgaagatttgaagggccagtgagcagatttattatgaaaaacagacaggggtacagg 420  
 QY 421 acagccagctcatcattgtcttgaactgaggaactccatgaagatctctttcttat 480  
 Db 421 acagccagctcatcattgtcttgaactgaggaactccatgaagatctctttcttat 480  
 QY 481 tcagagagggagggttaataggtctcagagatcttgggtgcaattgttactgtgtgtgtg 540  
 Db 481 tcagagagggagggttaataggtctcagagatcttgggtgcaattgttactgtgtgtgtg 540  
 QY 541 aaagatttcaatgagacacagctggcccgaggttcggagacagtaagagatcatgtttccc 600  
 Db 541 aaagatttcaatgagacacagctggcccgaggttcggagacagtaagagatcatgtttccc 600  
 QY 601 gtgaatgacgcttctcaggtctcgaagcctcatccactcaattttgaaagaagctctgc 660  
 Db 601 gtgaatgacgcttctcaggtctcgaagcctcatccactcaattttgaaagaagctctgc 660  
 QY 661 atcgaaattcttagcagctgaaccatccaccatatgtgcaggagagatcatttcaagtgtc 720  
 Db 661 atcgaaattcttagcagctgaaccatccaccatatgtgcaggagagatcatttcaagtgtc 720  
 QY 721 gtgagagaaacggtctccgacatgcccgaacgtggaaggggtccctctcagcttcaag 780  
 Db 721 gtgagagaaacggtctccgacatgcccgaacgtggaaggggtccctctcagcttcaag 780  
 QY 781 atcaatgactcgtgcacactcaatgagaagcccttttctgtggaagacacttattactg 840  
 Db 781 atcaatgactcgtgcacactcaatgagaagcccttttctgtggaagacacttattactg 840  
 QY 841 tgcacagcgcctattttaaagaagttggcatgaaagtgcactccaggtcagcatgaac 900  
 Db 841 tgcacagcgcctattttaaagaagttggcatgaaagtgcactccaggtcagcatgaac 900  
 QY 901 gatggcctcttttattctcaggttctgtcatcatccaccacacactgt 951  
 Db 901 gatggcctcttttattctcaggttctgtcatcatccaccacacactgt 951

RESULT 9  
 US-09-223-546-10  
 ; Sequence 10, Application US/09223546  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Holtzman, Douglas  
 ; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM  
 ; FILE REFERENCE: 09404/066001  
 ; CURRENT APPLICATION NUMBER: US/09/223,546  
 ; CURRENT FILING DATE: 1998-12-30  
 ; NUMBER OF SEQ ID NOS: 22  
 ; SOFTWARE: FastSeq for Windows Version 3.0

[illegible]

Db 661 atcgaaattctagcagctgaaccatccaccatattgtcaggagagtcatttcaagtgtc 720  
Qy 721 gtgagagaaacggtctccgacatcccgcaacgtgacaggggtcctctgcagcttcaag 780  
Db 721 gtgagagaaacggtctccgacatcccgcaacgtgacaggggtcctctgcagcttcaag 780  
Qy 781 atcaatgactcgggtcacactcaatgagaagcccttttctgtggaagacacttattactg 840  
Db 781 atcaatgactcgggtcacactcaatgagaagcccttttctgtggaagatacttattactg 840  
Qy 841 tgcacgagcctattctaaagaagtgtgcatgaaagtgcactccaggtgcagcatgaac 900  
Db 841 tgcacgagcctattctaaagaagtgtgcatgaaagtgcactccaggtgcagcatgaac 900  
Qy 901 gatggcctcttttctccagttctgtcatcatcaccaccacacactgt 951  
Db 901 gatggcctcttttctccagttctgtcatcatcaccaccacacactgt 951

## RESULT 11

US-09-599-596-10  
; Sequence 10, Application US/09599596  
; GENERAL INFORMATION:  
; APPLICANT: Holtzman, Douglas  
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM  
; FILE REFERENCE: 09404/066001  
; CURRENT APPLICATION NUMBER: US/09/599,596  
; CURRENT FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 09/423,546  
; PRIOR FILING DATE: 1998-12-30  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 10  
; LENGTH: 999  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-599-596-10

Query Match 86.0%; Score 949.4; DB 22; Length 999;  
Best Local Similarity 99.9%; Pred. No. 5e-249;  
Matches 950; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atggccacgagcgagcgagagccctcgcatcggtctccagtggtctctttggccact 60  
Db 1 atggccacgagcgagcgagagccctcgcatcggtctccagtggtctctttggccact 60  
Qy 61 ctggtctcatctcgcgcgggcaagggggacgcagggaggtgggttccagcctgctac 120  
Db 61 ctggtctcatctcgcgcgggcaagggggacgcagggaggtgggttccagcctgctac 120  
Qy 121 ggcggattgacctgtacttcatctattttggacaaatcaggaaagtgtgctgcacactggaat 180  
Db 121 ggcggattgacctgtacttcatctattttggacaaatcaggaaagtgtgctgcacactggaat 180  
Qy 181 gaaatctattactttgtggaacagttggtccacaaatcaccagccacagttgagaatg 240  
Db 181 gaaatctattactttgtggaacagttggtccacaaatcaccagccacagttgagaatg 240  
Qy 241 tcccttattgttttctccaccggaggaacaccttaattgaactgacagaagacagagaa 300  
Db 241 tcccttattgttttctccaccggaggaacaccttaattgaactgacagaagacagagaa 300  
Qy 301 caaatcgtcaagccttagaagaactccagaaagtctgcaggagagacacttacatg 360  
Db 301 caaatcgtcaagccttagaagaactccagaaagtctgcaggagagacacttacatg 360  
Qy 361 catgaagatttgaagggccagtgagcagatttattatgaaaacagacaagggtacagg 420  
Db 361 catgaagatttgaagggccagtgagcagatttattatgaaaacagacaagggtacagg 420  
Qy 421 acagccagcgtcatcattgctttgactgatggagaactccatgaagtctcttttctat 480  
Db 421 acagccagcgtcatcattgctttgactgatggagaactccatgaagtctcttttctat 480

Db 421 acagccagcgtcatcattgctttgactgatggagaactccatgaagtctcttttctat 480  
Qy 481 tcagagagggaggttaataaggtctcagagatcttggtcaattggtttactgtgtgggtg 540  
Db 481 tcagagagggaggttaataaggtctcagagatcttggtcaattggtttactgtgtgggtg 540  
Qy 541 aaagatttcaatgagacacagctggcccggtattgcgacagtaaggatcatgtgtttccc 600  
Db 541 aaagatttcaatgagacacagctggcccggtattgcgacagtaaggatcatgtgtttccc 600  
Qy 601 gtgaatgacggctttcagggtctcgaaggcatcaccactcaattttgaaagaagtcctgc 660  
Db 601 gtgaatgacggctttcagggtctcgaaggcatcaccactcaattttgaaagaagtcctgc 660  
Qy 661 atcgaaattcttagcagctgaaccatccaccatattgtcagagagatcatttcaagtgtc 720  
Db 661 atcgaaattcttagcagctgaaccatccaccatattgtcagagagatcatttcaagtgtc 720  
Qy 721 gtgagagaaacggtctccgacatcccgcaacgtggacaggggtcctctgcagcttcaag 780  
Db 721 gtgagagaaacggtctccgacatcccgcaacgtggacaggggtcctctgcagcttcaag 780  
Qy 781 atcaatgactcgggtcacactcaatgagaagcccttttctgtggaagacacttattactg 840  
Db 781 atcaatgactcgggtcacactcaatgagaagcccttttctgtggaagatacttattactg 840  
Qy 841 tgtccagcgcctattctaaagaagtgtggcatgaaagtgcactccaggtcagcatgaac 900  
Db 841 tgtccagcgcctattctaaagaagtgtggcatgaaagtgcactccaggtcagcatgaac 900  
Qy 901 gatggcctcttttctccagttctgtcatcatcaccaccacacactgt 951  
Db 901 gatggcctcttttctccagttctgtcatcatcaccaccacacactgt 951

## RESULT 12

US-09-970-076-7  
; Sequence 7, Application US/09970076  
; GENERAL INFORMATION:  
; APPLICANT: Young, John A.T.  
; APPLICANT: Bradley, Kenneth A.  
; APPLICANT: Collier, Robert J.  
; APPLICANT: Mogridge, Jeremy S.  
; TITLE OF INVENTION: Anthrax Toxin Receptor  
; FILE REFERENCE: 960296.97745  
; CURRENT APPLICATION NUMBER: US/09/970,076  
; CURRENT FILING DATE: 2001-10-03  
; PRIOR APPLICATION NUMBER: 60/251,481  
; PRIOR FILING DATE: 2000-12-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 2112  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (113)..(1111)  
US-09-970-076-7

Query Match 86.0%; Score 949.4; DB 36; Length 2112;  
Best Local Similarity 99.9%; Pred. No. 6.8e-249;  
Matches 950; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atggccacgagcgagcgagagccctcgcatcggtctccagtggtctctttggccact 60  
Db 113 atggccacgagcgagcgagagccctcgcatcggtctccagtggtctctttggccact 172  
Qy 61 ctggtctcatctcgcgcgggcaagggggacgcagggaggtgggggtccagcctgctac 120  
Db 173 ctggtctcatctcgcgcgggcaagggggacgcagggaggtgggggtccagcctgctac 232



Db 1113 gatggcctctcttttatctccagttctgtcatcatcaccaccacactgt 1163  
|||||

## RESULT 14

US-09-223-546-8  
; Sequence 8, Application US/09223546  
; GENERAL INFORMATION:  
; APPLICANT: Holtzman, Douglas  
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM  
; FILE REFERENCE: 09404/066001  
; CURRENT APPLICATION NUMBER: US/09/223,546  
; CURRENT FILING DATE: 1998-12-30  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 8  
; LENGTH: 2272  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (213)...(1211)  
US-09-223-546-8

Query Match 86.0%; Score 949.4; DB 16; Length 2272;  
Best Local Similarity 99.9%; Pred. No. 7e-249;  
Matches 950; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atggccacggcgagcgagagccctcgcatcgcttcacagtgctctctttggccact 60  
Db 213 atggccacggcgagcgagagccctcgcatcgcttcacagtgctctctttggccact 272  
Qy 61 ctggtgctcatctgcgcgggcaaggagggagcaggagagtgagggtccagcctgctac 120  
Db 273 ctggtgctcatctgcgcgggcaaggagggagcaggagagtgagggtccagcctgctac 332  
Qy 121 ggcgatttgacctactctatttggacaaatcagaaagtgtctgcaccactggaaat 180  
Db 333 ggcgatttgacctactctatttggacaaatcagaaagtgtctgcaccactggaaat 392  
Qy 181 gaaatctattactttgtggaacagttggttcacaaattcatcagccacagttgagaatg 240  
Db 393 gaaatctattactttgtggaacagttggttcacaaattcatcagccacagttgagaatg 452  
Qy 241 tcccttattgtttccaccggaggaacacacttaataatgaaactgacagaagacagaaa 300  
Db 453 tcccttattgttttccaccggaggaacacacttaataatgaaactgacagaagacagaaa 512  
Qy 301 caaatcccgtaaggcctagagaactcagaagtgtctgcaccactggaaat 180  
Db 513 caaatcccgtaaggcctagagaactcagaagtgtctgcaccactggaaat 392  
Qy 421 acagccagcgtcatctatgctttgactatggagaactccatgaagatctcttttctat 480  
Db 633 acagccagcgtcatctatgctttgactatggagaactccatgaagatctcttttctat 692  
Qy 481 tcagagagggaggtctaaggtctcgagatcttgggtgcaattgtttactgtgttggttg 540  
Db 693 tcagagagggaggtctaaggtctcgagatcttgggtgcaattgtttactgtgttggttg 752  
Qy 541 aaagatttcaatgacacagctgcccggattgctggcagtaagtaatgtgtttccc 600  
Db 753 aaagatttcaatgacacagctgcccggattgctggcagtaagtaatgtgtttccc 812  
Qy 601 gfgaatgacgggttcagggctgcaaggcatcatccactcaattttgaagaagtcctg 660  
Db 813 gfgaatgacgggttcagggctgcaaggcatcatccactcaattttgaagaagtcctg 872

Qy 661 atcgaataatctagcagctgaaccatccaccatatgtgcaggagagtcatttcaagtgtc 720  
Db 873 atcgaataatctagcagctgaaccatccaccatatgtgcaggagagtcatttcaagtgtc 932  
Qy 721 gtgagaggaacggcttcgcacatcccgcaacgtggacaggggtcctctgcagctccaag 780  
Db 933 gtgagaggaacggcttcgcacatcccgcaacgtggacaggggtcctctgcagctccaag 992  
Qy 781 atcaatgactcgggtcacactcaatgaagccctttctgtggaagacacttattactg 840  
Db 993 atcaatgactcgggtcacactcaatgaagccctttctgtggaagactattactg 1052  
Qy 841 tgtccagcgctatctttaaagaagtgtgcatgaaagtgcactccaggtcagcatgaac 900  
Db 1053 tgtccagcgctatctttaaagaagtgtgcatgaaagtgcactccaggtcagcatgaac 1112  
Qy 901 gatggcctctcttttatctccagttctgtcatcatcaccaccacactgt 951  
Db 1113 gatggcctctcttttatctccagttctgtcatcatcaccaccacactgt 1163  
|||||

## RESULT 15

US-09-471-179-8  
; Sequence 8, Application US/09471179  
; GENERAL INFORMATION:  
; APPLICANT: Holtzman, Douglas  
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM  
; FILE REFERENCE: 7853-173  
; CURRENT APPLICATION NUMBER: US/09/471,179  
; CURRENT FILING DATE: 1999-12-23  
; PRIOR APPLICATION NUMBER: 09/223,546  
; PRIOR FILING DATE: 1998-12-30  
; NUMBER OF SEQ ID NOS: 135  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 2272  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (213)...(1211)  
US-09-471-179-8

Query Match 86.0%; Score 949.4; DB 18; Length 2272;  
Best Local Similarity 99.9%; Pred. No. 7e-249;  
Matches 950; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atggccacggcgagcgagagccctcgcatcgcttcacagtgctctctttggccact 60  
Db 213 atggccacggcgagcgagagccctcgcatcgcttcacagtgctctctttggccact 272  
Qy 61 ctggtgctcatctgcgcgggcaaggagggagcaggagagtgagggtccagcctgctac 120  
Db 273 ctggtgctcatctgcgcgggcaaggagggagcaggagagtgagggtccagcctgctac 332  
Qy 121 ggcgatttgacctactctatttggacaaatcagaaagtgtctgcaccactggaaat 180  
Db 333 ggcgatttgacctactctatttggacaaatcagaaagtgtctgcaccactggaaat 392  
Qy 181 gaaatctattactttgtggaacagttggttcacaaattcatcagccacagttgagaatg 240  
Db 393 gaaatctattactttgtggaacagttggttcacaaattcatcagccacagttgagaatg 452  
Qy 241 tcccttattgttttccaccggaggaacacacttaataatgaaactgacagaagacagaaa 300  
Db 453 tcccttattgttttccaccggaggaacacacttaataatgaaactgacagaagacagaaa 512  
Qy 301 caaatcccgtaaggcctagagaactcagaagtgtctgcaccactggaaat 360  
Db 513 caaatcccgtaaggcctagagaactcagaagtgtctgcaccactggaaat 572  
Qy 361 catgaaggatttgaaaggccagtgagcagattttatgaaaaacagaaaggttacagg 420  
Db 573 catgaaggatttgaaaggccagtgagcagattttatgaaaaacagaaaggttacagg 632  
Qy 421 acagccagcgtcatctatgctttgactatggagaactccatgaagatctcttttctat 480  
Db 633 acagccagcgtcatctatgctttgactatggagaactccatgaagatctcttttctat 692  
Qy 481 tcagagagggaggtctaaggtctcgagatcttgggtgcaattgtttactgtgttggttg 540  
Db 693 tcagagagggaggtctaaggtctcgagatcttgggtgcaattgtttactgtgttggttg 752  
Qy 541 aaagatttcaatgacacagctgcccggattgctggcagtaagtaatgtgtttccc 600  
Db 753 aaagatttcaatgacacagctgcccggattgctggcagtaagtaatgtgtttccc 812  
Qy 601 gfgaatgacgggttcagggctgcaaggcatcatccactcaattttgaagaagtcctg 660  
Db 813 gfgaatgacgggttcagggctgcaaggcatcatccactcaattttgaagaagtcctg 872

|||||  
Db 573 catgaagatttgaaggccagtgagcagatttattatgaacacagaggggtacagg 632  
Qy 421 acagccagcgtcatctgttctgactgtgagaaacctccatgaagatctcttttctat 480  
Db 633 acagccagcgtcatctgttctgactgtgagaaacctccatgaagatctcttttctat 692  
Qy 481 tcagagaggaggctaataggctcagatcttgggtgcaattgtttactgtgtgtgtg 540  
Db 693 tcagagaggaggctaataggctcagatcttgggtgcaattgtttactgtgtgtgtg 752  
Qy 541 aaagatttcaatgagacacagcgtgcccggattgcgacagtaagagatcatgtttccc 600  
Db 753 aaagatttcaatgagacacagcgtgcccggattgcgacagtaagagatcatgtttccc 812  
Qy 601 gtgaatgacggttctcaggtctctgcaaggcatcatccactcaattttgaagaagtcctgc 660  
Db 813 gtgaatgacggttctcaggtctctgcaaggcatcatccactcaattttgaagaagtcctgc 872  
Qy 661 atcgaaattctagcagctgaacctccaccatattgtgcaggagagtcattttcaagttgtc 720  
Db 873 atcgaaattctagcagctgaacctccaccatattgtgcaggagagtcattttcaagttgtc 932  
Qy 721 gtgagagaaacggtctccgacatgcccgaacgtgacaggtcctctgcagcttcaag 780  
Db 933 gtgagagaaacggtctccgacatgcccgaacgtgacaggtcctctgcagcttcaag 992  
Qy 781 atcaatgactcgtcacactcaatgagaagcccttttctgtggaagacacttatttactg 840  
Db 993 atcaatgactcgtcacactcaatgagaagcccttttctgtggaagacacttatttactg 1052  
Qy 841 tgtccagcgcctatcttaaaagaagttggcatgaaagctgcactccaggtcagcatgaac 900  
Db 1053 tgtccagcgcctatcttaaaagaagttggcatgaaagctgcactccaggtcagcatgaac 1112  
Qy 901 gatggcctctctttatctccagttctgtcatcaccaccacacactgt 951  
Db 1113 gatggcctctctttatctccagttctgtcatcaccaccacacactgt 1163

Search completed: August 9, 2002, 01:54:14  
Job time: 12956 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 9, 2002, 00:23:01 ; Search time 109.68 seconds  
(without alignments)  
2472.460 Million cell updates/sec

Title: us-09-970-076-1\_copy\_104\_1207  
Perfect score: 1104  
Sequence: 1 atggccacgcggcgag.....gtgaggaataataaaaa 1104

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs; 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	4.3	7218	1	US-08-232-463-14
2	41.4	3.8	7218	1	Sequence 14, Appl
3	37.4	3.4	3489	2	US-08-232-463-14
4	37.4	3.4	32207	2	US-08-728-323A-1
5	37.4	3.4	32207	4	US-08-770-379-20
6	37.4	3.4	32207	4	US-08-757-669A-20
7	37.2	3.4	488	4	US-09-230-371A-20
8	35.2	3.2	1440	1	US-09-385-982-368
9	35.2	3.2	1440	1	US-07-882-202A-3
10	35.2	3.2	1440	1	US-08-021-615A-3
11	35.2	3.2	1440	3	US-08-321-777-3
12	35.2	3.2	1440	3	US-09-009-217-13
13	35.2	3.2	1440	5	US-09-009-656-13
14	35.2	3.2	2422	1	PCT-US93-04493-3
15	35.2	3.2	2422	1	US-08-475-845-1
16	35.2	3.2	2422	1	US-08-327-690-1
17	35.2	3.2	2422	2	US-08-660-289-1
18	35.2	3.2	2422	2	US-08-537-807-1
19	35.2	3.2	2422	3	US-08-871-003-1
20	35.2	3.2	2422	3	US-08-464-233-1
21	35.2	3.2	2422	4	US-09-189-607-1
22	35.2	3.2	2422	4	US-09-378-907-1
23	35.2	3.2	2422	5	PCT-US94-05779-1
24	35.2	3.2	2462	3	US-08-479-733A-25
25	35.2	3.2	2462	3	US-08-487-427-25
26	35.2	3.2	2462	3	US-08-479-727A-25
27	35.2	3.2	2462	5	US-08-482-369A-25
					PCT-US95-07439-25

C 28 34.6 3.1 2252 4 US-08-462-467B-13 Sequence 13, Appl  
C 29 34.6 3.1 2252 4 US-08-334-179A-13 Sequence 13, Appl  
C 30 33.2 3.0 192 2 US-08-700-637-8 Sequence 8, Appl  
C 31 33 3.0 230 2 US-08-700-637-10 Sequence 10, Appl  
C 32 33 3.0 2881 2 US-08-570-227A-1 Sequence 1, Appl  
C 33 33 3.0 2881 4 US-09-077-991-1 Sequence 1, Appl  
C 34 32.6 3.0 273 4 US-09-318-978-14 Sequence 14, Appl  
C 35 32.6 3.0 289 4 US-09-007-005-17 Sequence 17, Appl  
C 36 32.6 3.0 289 4 US-09-244-796-17 Sequence 17, Appl  
C 37 32.4 2.9 1120 1 US-08-203-806B-3 Sequence 3, Appl  
C 38 32.4 2.9 1120 4 US-09-017-754A-3 Sequence 3, Appl  
C 39 32.4 2.9 2580 3 US-09-050-863-2 Sequence 2, Appl  
C 40 32.4 2.9 2580 4 US-09-359-081-2 Sequence 2, Appl  
C 41 32.4 2.9 5452 2 US-09-130-114-1 Sequence 1, Appl  
C 42 32.4 2.9 9600 4 US-08-910-647-1 Sequence 1, Appl  
C 43 32.4 2.9 10596 1 US-07-884-811-15 Sequence 15, Appl  
C 44 32.4 2.9 10596 1 US-07-885-971-15 Sequence 15, Appl  
C 45 32.4 2.9 10596 1 US-08-087-783A-15 Sequence 15, Appl

## ALIGNMENTS

RESULT 1  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-Fls  
US-08-232-463-14

Query Match 4.3%; Score 47; DB 1; Length 7218;  
Best Local Similarity 6.7%; Pred. No. 0.00014;  
Matches 26; Conservative 199; Mismatches 164; Indels 0; Gaps 0;  
QY 689 ccattatgtcagagagtcattcaagtgtcgtgagaggaacacggtctccgacatccc 748  
Db 1011 CCATACGCTACAGAAATAATCCGAGCTGGCTGAGGTCGAGGAGCTGCGATYYY 1070  
QY 749 gcaacgtggacaggtctctgcagcttcaagatcaatgactcgtgtcacatcaatgaga 808  
Db 1071 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1130  
QY 809 agccctttctgtggaagacactattactgtgtccagcctctctctaaagaagtgg 868  
Db 1131 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1190  
QY 869 gcatgaagctgcactcagctcagctgaacgatgctctctctttatctccagttctg 928  
Db 1191 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1250  
QY 929 teatcatcacacacacactgtctgacggttccatctcctgacatgcctgctatcc 988  
Db 1251 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1310  
QY 989 ttctctctctagcctgctctctctctgtgtgtctgctcctctgctgacactgtga 1048  
Db 1311 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1370  
QY 1049 ttatcaagaggtctccctcacccctgcc 1077  
Db 1371 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1399

RESULT 2  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: PTZgpt-F1s  
; US-08-232-463-14  
Query Match 3.8%; Score 41.4; DB 1; Length 7218;  
Best Local Similarity 6.8%; Pred. No. 0.0094;  
Matches 27; Conservative 198; Mismatches 174; Indels 0; Gaps 0;  
QY 263 gaggaacaaccttaataactgacagagacagagaacaaatccgtcaaggcctagaag 322  
Db 1407 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1348  
QY 323 aactcccgaaagtctcgcaggaggagacacttacatgcatgagagattgaaaggcca 382  
Db 1347 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1288  
QY 383 gtgacgacttattatgaaacagacagaggtacagagcagcagcgctcatctgctt 442  
Db 1287 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1228  
QY 443 tgactgatgagaaactccatcagatctctttcttattcagagagggaggttaagt 502  
Db 1227 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1168  
QY 503 ctgagatcttggtgcaattgtttactgtgtgtgtgaaagatttcaatgagacagc 562  
Db 1167 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1108  
QY 563 tggcccgattcgcgacagtaagatcatgtgttcccgtaagtacggtccttcaggctc 622  
Db 1107 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1048  
QY 623 tgcaaggcctcatccatccatcttattgaaagactcgtca 661  
Db 1047 TGCAGCCAAAGCTCGGAATTAATCTGTGAGCGTATGGCA 1009

RESULT 3  
US-08-728-323A-1/c  
; Sequence 1, Application US/08728323A  
; Patent No. 5948676  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Yuan  
; APPLICANT: Bohenzky, Roy A.  
; APPLICANT: Russo, James J.  
; APPLICANT: Edelman, Isidore S.  
; APPLICANT: Moore, Patrick S.  
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's  
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA  
; TITLE OF INVENTION: Encoding Same And Uses Thereof  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/728,323A  
; FILING DATE:  
; CLASSIFICATION: 435





```

;
; LENGTH: 32207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
US-08-757-669A-20

Query Match          3.4%; Score 37.4; DB 4; Length 32207;
Best Local Similarity 50.3%; Pred. No. 0.52;
Matches 92; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 883 ctccagtgacgatgaacgatggccctctcttttatctccagttctgtcatcaccacc 942
    |||| |||| || || || || || || || || || || || || || || || || ||
Db 20848 CTCTCTGTCATCTCTCTGTCATCTCTCTGTCATCTCTCTGTCATCTCTCTGTCATC 20907

QY 943 acacactgttctgaagttccatctcctggccatcgccctgctgctgctctctctccta 1002
    || || || || || || || || || || || || || || || || || || || || ||
Db 20908 CTCCTGTCATCTCTCTGTCATCTCTCTGTCATCTCTCTGTCATCTCTCTGTCATC 20967

QY 1003 gccctggctctctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1062
    |||| |||| || || || || || || || || || || || || || || || || ||
Db 20968 CTCCTGTCATCTCTCTGTCATCTCTCTGTCATCTCTCTGTCATCTCTCTGTCATC 21027

QY 1063 cct 1065
    ||
Db 21028 CTT 21030

RESULT 6
US-09-230-371A-20
; Sequence 20, Application US/09230371A
; Patent No. 6348586
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A
; APPLICANT: Russo, James J
; APPLICANT: Edelman, Isidore S
; APPLICANT: Moore, Patrick S
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
; APPLICATION NUMBER: 45185-G-PCT-US
; FILE REFERENCE: 45185-G-PCT-US
; CURRENT FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: PCT/US97/13346
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 32207
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
;
US-09-230-371A-20

Query Match          3.4%; Score 37.4; DB 4; Length 32207;
Best Local Similarity 50.3%; Pred. No. 0.52;
Matches 92; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 883 ctccagtgacgatgaacgatggccctctcttttatctccagttctgtcatcaccacc 942
    |||| |||| || || || || || || || || || || || || || || || || ||
Db 20848 CTCCTGTCATCTCTCTGTCATCTCTCTGTCATCTCTCTGTCATCTCTCTGTCATC 20907

QY 943 acacactgttctgaagttccatctcctggccatcgccctgctgctgctctctccta 1002
    || || || || || || || || || || || || || || || || || || || || ||
Db 20908 CTCCTGTCATCTCTCTGTCATCTCTCTGTCATCTCTCTGTCATCTCTCTGTCATC 20967

QY 1003 gccctggctctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1062
    |||| |||| || || || || || || || || || || || || || || || || ||
Db 20968 CTCCTGTCATCTCTCTGTCATCTCTCTGTCATCTCTCTGTCATCTCTCTGTCATC 21027

QY 1063 cct 1065
    ||
Db 21028 CTT 21030

US-09-385-982-368
; Sequence 368, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDSER, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 368
; LENGTH: 488
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(488)
; OTHER INFORMATION: n = A,T,C or G
;
US-09-385-982-368

Query Match          3.4%; Score 37.2; DB 4; Length 488;
Best Local Similarity 50.8%; Pred. No. 0.035;
Matches 66; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 208 gctcacaaattcatcagccacagttgagaatgctcttttattgttttctccaccgagga 267
    |||| |||| || || || || || || || || || || || || || || || || ||
Db 268 gctcangaatcctgtt9gctggaagtcacagagtgctctttctgattctcctgaagatanatna 327

QY 268 acaactttaataactgcagagacagacagacagacacaaatcctcgaagcctagaagaaccc 327
    |||| |||| || || || || || || || || || || || || || || || || ||
Db 328 acagcccnngcccaangagagagnnntagtaacaagccnctcgcgtacctgtcncgg 387

QY 328 cagaaagttc 337
    ||||
Db 388 gcggnngttc 397

RESULT 8
US-07-882-202A-3
; Sequence 3, Application US/07882202A
; Patent No. 5374617
; GENERAL INFORMATION:
; APPLICANT: Morrissey, James H.
; APPLICANT: Comp, Philip C.
; TITLE OF INVENTION: Treatment of Bleeding with Modified
; TITLE OF INVENTION: Tissue Factor in Combination with FVIIa
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richards, Medlock & Andrews
; STREET: 1201 Elm Street, Suite 4500
; CITY: Dallas
; STATE: Texas
; COUNTRY: US
; ZIP: 75270-2197
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;

```







```

;
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: N
; ANTI-SENSE: N
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 28..1420 /codon_start= 28
; OTHER INFORMATION: /product= "Factor VII"
;
; US-08-475-845-1

```

```

Query Match 3.2%; Score 35.2; DB 1; Length 2422;
Best Local Similarity 62.5%; Pred. No. 0.46;
Matches 55; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 928 gtcatcaccacacacactgttctgacgtttccatctctggccatcgccctgctgac 987
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 836 GTCATCATCCCCAGCAGGTACGTCCCGGGCACCACCAACCAACGACATCGCGTGTCTCCGC 895

QY 988 ctgttctgctctagccctggctctcc 1015
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 896 CTGCACCAGCCCGTGGTCTCACTGACC 923

```

Search completed: August 9, 2002, 00:23:44  
Job time: 7997 sec

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: N
; ANTI-SENSE: N
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 28..1420 /codon_start= 28
; OTHER INFORMATION: /product= "Factor VII"
;
; US-08-475-845-1

```

```

Query Match 3.2%; Score 35.2; DB 1; Length 2422;
Best Local Similarity 62.5%; Pred. No. 0.46;
Matches 55; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 928 gtcatcaccacacacactgttctgacgtttccatctctggccatcgccctgctgac 987
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 836 GTCATCATCCCCAGCAGGTACGTCCCGGGCACCACCAACCAACGACATCGCGTGTCTCCGC 895

QY 988 ctgttctgctctagccctggctctcc 1015
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 896 CTGCACCAGCCCGTGGTCTCACTGACC 923

```

RESULT 15  
US-08-327-690-1  
; Sequence 1, Application US/08327690  
; Patent No. 5817788  
; GENERAL INFORMATION:  
; APPLICANT: Berkner, Kathleen L.  
; APPLICANT: Petersen, Lars C.  
; APPLICANT: Hart, Charles E.  
; APPLICANT: Hedner, Ulla  
; APPLICANT: Bregengaard, Claus  
; TITLE OF INVENTION: Modified Factor VII  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Khourie and Crew  
; STREET: One Market Plaza, Steuart Street Tower  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 94105-1492  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/327,690  
; FILING DATE: 24-OCT-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/065,725  
; FILING DATE: 21-MAY-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/662,920  
; FILING DATE: 28-FEB-1991  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parmelee, Steven W.  
; REGISTRATION NUMBER: 31,990  
; REFERENCE/DOCKET NUMBER: 13952-8-3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-467-9600  
; TELEFAX: 415-543-5043  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2422 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single

---

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 8, 2002, 23:02:53 ; Search time 2803.4 seconds

(without alignments)

5315.202 Million cell updates/sec

Title: US-09-970-076-1\_COPY\_104\_1207

Perfect score: 1104

Sequence: 1 atggccacgagcgagcgag.....gtgaggaaataaaaaa 1104

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estinu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_gss:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	691.8	62.7	1063	9	AL542724
2	684.6	62.0	878	9	BI823853
3	681.4	61.7	1514	11	AK013005
4	637.8	57.8	725	10	BE741333
5	585	53.0	963	10	BE741333
6	582.6	52.8	820	10	BE741333
7	482	43.7	964	10	BE281831
8	447.4	40.5	638	10	BE369415
9	420.4	38.1	513	9	BE146084
10	420.4	38.1	542	9	BE146075
11	399.8	36.2	465	9	BE145958
12	384.8	34.9	487	9	BE146355
13	381.2	34.5	539	9	AI466645
14	376.4	34.1	693	10	BE323538
15	373.6	33.8	554	9	AI084806
16	329.8	29.9	356	9	BE146135
17	329	29.8	569	10	BM050390

18	304.6	27.6	608	10	BI391348
19	295.4	26.8	369	9	AA004013
20	277.6	25.1	464	9	AUI29171
21	274	24.8	767	9	AI905553
22	266.6	24.1	527	9	AV667876
23	246.8	22.4	470	9	AA037321
24	242.4	22.0	842	10	BG576342
25	241.8	21.9	287	10	BG999787
26	239.4	21.7	343	9	AI750523
27	227.6	20.6	389	10	BF833348
28	224.8	20.4	441	10	BI183103
29	220.8	20.0	664	10	BM423073
30	197	17.8	282	9	BE146276
31	195	17.1	430	10	BE842906
32	188.8	17.1	604	9	BB620906
33	175.6	15.9	672	10	BI869089
34	175.4	15.9	360	10	BE478445
35	168.6	15.3	434	9	AV749755
36	141.6	12.8	757	9	BB649806
37	137.2	12.4	562	9	AW106196
38	121.8	11.0	581	10	BM149748
39	121.6	11.0	723	10	BI917586
40	121.2	11.0	301	10	BG951999
41	120.8	10.9	985	10	BF531402
42	119.6	10.8	503	10	BE371536
43	118.4	10.7	582	10	BM150243
44	116.8	10.6	559	9	AV599556
45	113	10.2	811	9	AUI39036

## ALIGNMENTS

RESULT 1

AL542724

LOCUS

DEFINITION

AL542724 LTI\_FL002\_PL1

1063 bp mRNA linear EST 16-FEB-2001

AL542724

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

AL542724 LTI\_FL002\_PL1 Homo sapiens cDNA clone CS0DE011YL09 5 prime mRNA linear EST 16-FEB-2001  
1063 bp  
mRNA sequence.  
AL542724.1 GI:12875049  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
Location/Qualifiers  
1. .1063  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DE011YL09"  
/lab\_host="LTI\_FL002\_PL1"  
/note="Organ: Placenta; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville , Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT

ORIGIN

272 a 254 c 314 g 221 t

2 others





cdNA Library Preparation: Ling Hong/Rubin Laboratory  
cdNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LLC810 row: n column: 19  
High quality sequence stop: 711.  
Location/Qualifiers  
1. 725  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3948114"  
/clone\_lib="NIH\_MGC\_9"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: ovary; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 206 a 165 c 176 g 178 t  
ORIGIN  
Query Match 57.8%; Score 637.8; DB 10; Length 725;  
Best Local Similarity 98.8%; Pred. No. 1.7e-163;  
Matches 653; Conservative 0; Mismatches 7; Indels 1; Gaps 1;  
QY 292 gacagagaacaatccgtcaagccttagaagcctcagaactccagaagtctgcagagagagac 351  
Db 1 GACAGAGAACAATCCGTCAGGCCCTAGAGAACTCCAGAAAGTTCGCCAGGAGGAGAC 60  
QY 352 acttacatcatgaagatttgaaggccagtcagagcagtcagagcagatttattatgaacagacaa 411  
Db 61 ACTTACATCATGAAGATTGTAAGGGCCAGTCAGCAGATTTATTATGAAACAGACAA 120  
QY 412 ggtacagacagcagcagtcattgttcagtcagtcagtcagtcagtcagtcagtcagtcagtc 471  
Db 121 GGGTACAGACACGCCAGCGTCATTCATTGCTTTCAGTGTGAGAACTCCATGAAGATCTC 180  
QY 472 tttttcttcagagagggaggtcctaagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 531  
Db 181 TTTTTCATTACAGAGGGAGGCTAATAGTCTCGAGATCTTGTGTCATTTGTTTACTGT 240  
QY 532 gttggtgtgaaagatttcaatgagacacagctggcccgagtcgagcagtcagtcagtcagtcagtc 591  
Db 241 GTTGGTGTGAAAGATTTCATGAGACACACAGCTGGCCGGATTGGGACACAGATGATCAT 300  
QY 592 gtgtttccctgaatgacgcttcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 651  
Db 301 GTGTTTCCCGTGAATGACGGCTTTCAGGCTCTGCAAGGCATCATCCACTCAATTTTGAAG 360  
QY 652 aagtcctgcatacgaattcttagcagctgaaccatccaccatattgtgagagagtcattt 711  
Db 361 AAGTCTCTGCATCGAAATTCCTAGCAGCTGAACCATCCACCATATGTGCAGAGAGATCAATT 420  
QY 712 caagtgtctgagagaaacgcttcagacatgcccgcaacgctgagcagtcagtcagtcagtcagtc 771  
Db 421 CAAGTTGTCTGAGAGGAAACGGCTTCCGACATGCCCGCAACGTCGAGAGGGTCTCTCTGC 480  
QY 772 agctcaagatcaatgactcggtcagtcacactcaatgagagcccttttctgtggaagacact 831  
Db 481 AGCTTCAGATCAATGACTCGTGCACACTCAATGAGAAAGCCCTTTTCTGTGGAAGATACT 540  
QY 832 tattactgtgtccagcgcctatttaa-aagaagtgtgcatgaaagcactccaggt 890  
Db 541 TATTACTGTCTCCAGGCCCTATCTTAACAAGAAGTTGGCATGGAAGCTGCACCTCCAGGT 600  
QY 891 cagatgaacatgacctctcttttatccagtcctgtcagtcagtcagtcagtcagtcagtcagtcagtc 950  
|||||

328 cagaagttctgcccaggaggagacacttacatgcatgaagatttgaagggccagtgag 387  
|||||  
Db 62 CAGAAAGTTCTCCAGGAGGAGACACTTACATGACGAGGATTCGAGAGGCCAGTGAG 121  
|||||  
QY 388 cagatttattatgaacacagacaaggtacagagcagcagcagtcattctgttgcact 447  
|||||  
Db 122 CAGATTACTATGAGAACAGTCAAGGATACAGGACGGCGGCTCATCTCGCTTGAGC 181  
|||||  
QY 448 gatggagaactccatgaagatctcttttctatttcagagagggaggttaataaggtctcga 507  
|||||  
Db 182 GATGGGAGCTGCAGGAGGACCTCTTCTTACTCAGAGAGGGAGGCTAAGCGATCCCGA 241  
|||||  
QY 508 gatctggtgcaattgtttactgtgtgtggaagatttcaatgagacacagctggcc 567  
|||||  
Db 242 GACCTTGGTGGCATGTTTACTGCGGTGGCGTGAAGGATTTCAATGAAGAACTCAGTTGGT 301  
|||||  
QY 568 cggattgcgacagtaagatcatgtgtttcccgtaagtcagcggctttcaggtctctgcaa 627  
|||||  
Db 302 CGGATTGCAGACAGTAAGGACACCGGTGTTCTGTGAACGACGGCTTCAGGCTCTCCAA 361  
|||||  
QY 628 ggcataccactcaattttgaaagagtcctgcatacgaaattctagcagctgaacatcc 687  
|||||  
Db 362 GGCATTATCCACTCAATTTTAAAGAAATCTGCATCGAAATTCGTGGCGCTGAACCATCC 421  
|||||  
QY 688 accatgtgcagagcagtcattcctcaagttgtcgtgagagaaacgcttcacgacatgcc 747  
|||||  
Db 422 ACCATCTGCGGGGAGAGTCTTTCAGTGGTGTGAAGGAAATGGCTTCGACATGCC 481  
|||||  
QY 748 cgcaagtgacaggggtctctcagctcagtcacatgaatgaactcagtcagtcagtcagtcagtc 807  
Db 482 CGCAATGTGGACAGGCTCTCTGCAGCTTCAAAATCAATGACTCAGTCAGCTCAATGAG 541  
|||||  
QY 808 aagccctttctgtggaagacacttattactgtgtccagcgcctattcttaaaagaagt 867  
|||||  
Db 542 AAGCCCTTTCGTGGGAAGACATTATTGCTGTGCCACGACCAACATCTTGAAGAAGTT 601  
|||||  
QY 868 ggcataaagctgcaactccaggtcagcatgaacgatggcctctcttttatccagttct 927  
Db 602 GGCATGAAGCTGCATGTCAGGTCAGCATGACGACGGCTGCTTCTTCATCTCCAGTTCT 661  
|||||  
QY 928 gtcatcatcaccacacacactgtctcagcaggttccatcctgcccagtcgcccagtcgac 987  
Db 662 GTCATCATCACACACACACTGTTCAGACGGCTCCATCTCGGCGCTGCTCTGCTGGTC 721  
|||||  
QY 988 ctgtctcgtctcagccctggctcctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1047  
Db 722 CTCCTCTCTGCTGCGCCCTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 781  
|||||  
QY 1048 attatcaaggagggtccctccacccctgccagagagagagtgaggaaaa 1094  
Db 782 ATCATCAAGGAGGTCCTCCACCCCTGTTGAGGAGAGTGAGGAAGA 828  
|||||

RESULT 4  
BE741333  
LOCUS 601594179f1 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3948114 5',  
DEFINITION mRNA sequence.  
ACCESSION BE741333  
VERSION BE741333.1 GI:10155325  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 725)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-r@mail.nih.gov  
Tissue Procurement: DCTD/BTP

```
Db 601 CAGCATGAACGATGGCCTCTCTTTTATCTCCAGTTCTGTCTCATCATCACCAGCAACTTG 660
Qy 951 t 951
Db 661 T 661

RESULT 5
BG326444
LOCUS 602425082F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4563020 5',
DEFINITION mRNA sequence.
ACCESSION BG326444
VERSION BG326444.1 GI:13132881
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1276 row: c column: 21
High quality sequence stop: 718.
FEATURES
source
1..963
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4563020"
/clone_lib="NIH_MGC_14"
/tissue_type="renal cell adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 261 a 217 c 292 g 193 t

Query Match 53.0%; Score 585; DB 10; Length 963;  
Best Local Similarity 96.3%; Pred. No. 5.4e-149;  
Matches 621; Conservative 0; Mismatches 20; Indels 4; Gaps 2;

```
Qy 1 atggccacggcggagcgagccctcgcatcgcttcacagtggtctctctttggccact 60
Db 114 ATGGCCACGGCGGAGCGGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 173
Qy 61 ctgtgctcatctgcgcggcgaaggcgacgagggagatgggggtccagctgctac 120
Db 174 CTGTGCTCATCTGCGCCGGCAGGGGACGACGAGGAGATGGGGGTCCAGCTGCTCTCTTTGGCCACT 233
Qy 121 ggcggattgacctgtactctatttggacaaaatcagaagtgctgctcacaccactggaat 180
Db 234 GCGCGATTGACCTGTACTTCTATTGTCACAAATCAGGAAGTGTGCTGCACCACTGGAAT 293
Qy 181 gaaatctattactttgtggacagttggtcacaaaattcaccagccacagttgagaatg 240
Db 294 GAAATCTATTACTTTGTGGAACAGTTGGCTCACAAATTCATCAGCCCACTGAGAAATG 353
```

```
Qy 241 tcctttattgtttctccacccggaggaaacaccttaataaactgacagagagagaa 300
Db 354 TCCTTTATTGTTTCTCCACCCGAGGAACAACCTTAATGAACCTGACAGAGACAGAGAA 413
Qy 301 caaatccgctcaagcctagaaactccagaagaattcttccaggaggagacactacatg 360
Db 414 CAAATCCGTCGAAGCCCTAGAAAGAACTCCAGAAAGTTCTGCCAGGAGGAGACACTTACATG 473
Qy 361 catgaagatttgaagggccagtcagcagatttatatgaatacagaaggtacagg 420
Db 474 CATGAAGGATTTCGAAAGGGCCAGTGAGCAGATTTATTATGAAACACAGACAAGGGTACAGG 533
Qy 421 acagccagcgtcatctgttcttgactgatggagaactccatgaagatcctctttctcat 480
Db 534 ACAGCCAGCGTCATCTGCTTTCACCTGATGGAGAACTCCATGAAGATCTCTTTTCTAT 593
Qy 481 tcagagggggaggttaataagtcctcgagatcttgggtgcaattgttactggtggtggtg 540
Db 594 TCAGAGAGGGGAGGCTAATAGGTCTCGAGATCTTGGTGCATTTGTTACTGTTGTTGGTGTG 653
Qy 541 aaagatttcaatgagacacagctggcccgattcgagacagtaagatcatctatttccc 600
Db 654 AAAAGATCTCATGAGACAAAGTTTGCC--GGATCGGACAGTAGGATCATGTGTATCCC 711
Qy 601 gtgaatgacggcttccaggtctcgaagcctcatcctccactcaatt 645
Db 712 GTGAATGACGG--TTTAAAGATCTGGAAGGCATCATCCATTCATT 754

RESULT 6
BG281561
LOCUS 602402412F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4544691 5',
DEFINITION mRNA sequence.
ACCESSION BG281561
VERSION BG281561.1 GI:13030486
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM1228 row: h column: 04
High quality sequence stop: 815.
FEATURES
source
1..820
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 209 a 189 c 240 g 181 t

```

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCMI230 row: a column: 15
High quality sequence stop: 659.
Location/Qualifiers
1. .964
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4545302"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/notes="organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
237 a 241 c 283 g 203 t

```

ORIGIN	Query Match	52.8%;	Score 582.6;	DB 10;	Length 820;
	Best Local Similarity	96.1%;	Pred. No. 2.3e-148;		
	Matches 683;	Conservative	0;	Mismatches 19;	Indels 9; Gaps 8;
QY	1	atggccacggcgagcgagagccctcgccatcgctccagctccagctgctctctttggccact	60		
DB	112	ATGGCCACGGCGAGCGGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT	171		
QY	61	ctggtgctcatctggccgggcaaggggacgcaggggaggtgggggtccagctcgtctac	120		
DB	172	CTGCTGCTCATCTGCGCGCGGCAAGGGGACGCAGGGAGGATGGGGTCCAGCCTGCTTAC	231		
QY	121	ggcggatttgacctgtactctatcttttgacaatacaggaagtgtctccaccactgggaat	180		
DB	232	GGCGGATTTGACCTGTACTTCAATTTTGGACAATCAGGAAGTGTGCTGCACCACTGGAAAT	291		
QY	181	gaatctattacttttgaaacagtt-ggctcacaaattcatcagccacagcttgagaaat	239		
DB	292	GAATCTATTACTTTGTGGAACAGTTGGGTTCACAATTTTCATCAGCCACAGTTGAAAT	351		
QY	240	gtct--tbtatttttctccacccagagaaacacotttaataaagactgacaaacacaga	297		
DB	352	GTCTTTATATGTGTTTCTCCACCCGAGGAACAAGCTTAATGAATCTGACGAAGACAGA	411		
QY	298	gaacaaatccgtcaagccttagaagaactccagaagaagtctgccaggaggagacacttac	357		
DB	412	GAACAAATCCGTCAGGCCCTAGAAGAACTCCAGAAAGTTCTGCCAGGAGGAGACACTTAC	471		
QY	358	atgcatgaagatttgaagggccagtgagcagatttatt-atgaataacagacaaggtta	416		
DB	472	ATGCATGAAGATTGTAAGAGGCCAGTGCAGCAGATTCAATCATGAATAACAGACAAGGTA	531		
QY	417	caggacacgcagctcatctgtcttgcactgatggagaactccatgaagatctctttt	476		
DB	532	CAGGACACGCCAGCGTCATCATGCTTGACTGATGGAGAACTCCATGAAGATCTCTTGG	591		
QY	477	ctattcagagaggaggtctaattagggtctgagatcttgggtgcaattgtttactgtgttg	536		
DB	592	CTATTACAGAGAGGAGGCGTATAGGTCTCGAGATCTTGGTGCATTTGTTTACTGTGATGG	651		
QY	537	tgtgaa-agatttcaatgagacacagctggcccgagattggcgacagt-aaggatcatgtg	594		
DB	652	TGTGAACAGATTTCATGAGACACAGCTGGCCCGATTGGCGACAGTAAAGGATCATGTG	711		
QY	595	tttccgtggaatgacggcttccaggtctgcaaggcatccactcaattttgaaagag	654		
DB	712	TTTCCGCTGAATGACGG-TTTTCAGGCTCTG-AGGCATCATCCATCAATCTTGACGAG	769		
QY	655	tctcctcagaaattctagcagctgaaccatccaccatattgaggagag	705		
DB	770	TCTTGCTCGAAATTTCTAGCAGCTGAACCATACA-CATATGTCGAGGACAG	819		
RESULT	7				
LOCUS	BG281831		964 bp	mRNA	linear
DEFINITION	602403057F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4545302 5', mRNA sequence.				EST 21-FEB-2001
ACCESSION	BG281831				
VERSION	BG281831.1	GI:13030757			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 964)				
JOURNAL	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .				
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)				
	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				

DEFINITION 601220816F1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:3589875 5',  
mRNA sequence.  
ACCESSION BE369415  
VERSION BE369415.1 GI:9314778  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 638)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM8756 row: p column: 04  
High quality sequence stop: 573.  
Location/Qualifiers  
1. .638  
/organism="Mus musculus"  
/strain="C57BL/6J (f129)"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3589875"  
/clone\_lib="NCI\_CGAP\_Lu29"  
/tissue\_type="spontaneous tumor, metastatic to mammary."  
Stem cell origin.  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"  
BASE COUNT 173 a 148 c 173 g 144 t  
ORIGIN  
Query Match 40.5%; Score 447.4; DB 10; Length 638;  
Best Local Similarity 85.8%; Pred. No. 2.1e-111;  
Matches 543; Conservative 0; Mismatches 86; Indels 4; Gaps 4;  
QY 154 tcaggaagtgtctgcaccactggaatgaatctattcttcttggaacagttggtctaac 213  
DB 1 TCAGGAAGTGTCTGCACCACCTGGAATGAAATCTACTTCTCGTGAGCAGCTGCTCAT 60  
QY 214 aaattcatcagccacagtgagaatgctcttattgtttctccaccgaggaacaacc 273  
DB 61 AGATTTCATCAGCCACAGCTAAGATGTCCTTCATTGCTTCTCTACTCTCAGGAGCAACT 120  
QY 274 ttaataaactga-cagagacagagaaacaaatccg-tcaaggcctagaagaactccaga 331  
DB 121 TTAATGAACAACTAAGCTGAGGACAGGAACAGATCCGCAAGGCCTAGAGAGCTCCAGA 180  
QY 332 aagttctccagagagagacacttacatgcatgagagatttgaaaggccagtgagcaga 391  
DB 181 AAGTTCTGCCAGGGAGACACTTACATGTCAGAGGAGGATTTGAGAGGCCACCTAGCAGCA 240  
QY 392 ttattatgaacacagaaaggttacagacagccagcgtcatcattgtttgactgatg 451  
DB 241 TTTATATGAGACAGTCAAGATACAGGACAGCCAGCGTCATCATCGCTTCAGCGATG 300  
QY 452 gagaacccatgaagatctcttttctatctcagagagggagctaataggtctcagatc 511  
DB 301 GGGAGCTGCACGAGGACCTCTTCTTCTACTCAGAGAGGAGGCTAACAGATCCCGAGACC 360  
QY 512 ttggtgcaattgttactgtgtgtgtgaaagatttcaatgagacagctggcccgga 571  
DB 361 TTGGTGCAG-TGTTTACTCGTGTGGCGTGAAGGATTTCAATGAAACTCATGTTGGTCGGA 419

QY 572 ttgcgacagtaagatcatgtgtttccgtaagacggctttcaggctctcgaaggca 631  
DB 420 TTGCAGACAGTAAGGACACAGCTGTTTCTGTGACGACGGCTTCCAGGCTCTCCAAGGCA 479  
QY 632 tcatccactcaattttgaagaagctctcatcgatgaattctagcagctgaaccatcacca 691  
DB 480 TTATCCACTCAATTTTAAAGAAATCTTGTCATCGAAATTTCTGGCGCTGAACCATCCAGCA 539  
QY 692 tatgtcaggagagtcatttcaagttgtcgtgagagaaacggcttcgacatgccgca 751  
DB 540 TCTGCGCGGAGAGTCTTCAAGTGGTCGTAAAGAGAAATGGCTTCGACATGTCGCG- 598  
QY 752 acgtgacagaggtctctgcagcttcaagatca 784  
DB 599 ATGTGACACAGGCCCTGGGGTCTAAATAATGA 631  
RESULT 9  
BE146084/c  
LOCUS BE146084 513 bp mRNA linear EST 21-JUN-2000  
DEFINITION MRO-HT0209-280300-106-g09 HT0209 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BE146084  
VERSION BE146084.1 GI:8608808  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 513)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?t1=st2-MRO-HT0209-280300-106-g09&t3=2000-03-28&t4=1>)  
Seq primer: puc 18 forward  
High quality sequence start: 21  
High quality sequence stop: 513.  
Location/Qualifiers  
1. 513  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT0209"  
/dev\_stage="Adult"  
/note="Organ: head\_neck; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
BASE COUNT 108 a 150 c 127 g 128 t  
ORIGIN  
Query Match 38.1%; Score 420.4; DB 9; Length 513;

Best Local Similarity	99.8%	Pred. NO. 4.9e-104;			
Matches	421;	Conservative	0;	Mismatches	1;
				Indels	0;
QY	1	atggccacgaggagcagcctcgatcggttcacgtggctctcttggccact	60		
DB	426	ATGGCCACGGCGAGAGCCCTCGGCATCGGCTCCAGTGGCTCTCTTTGGCCACT	367		
QY	61	ctggtgtcatctgcgcgggcaaggaggggacgagggagatgggggtccacgctgctac	120		
DB	366	CTGGTGTCTCATCTGCGCGGCAAGGGGACGAGGGAGATGGGGGTCCAGCTGCTAC	307		
QY	121	gcgcatitgacctgactctatttggacaaatcagggaagtgtgtcgcaccactggaat	180		
DB	306	GCGGATTTGACCTGTACTTCTATTTTGGACAATCAGGAAGTGTGTCGCCACTTGAAT	247		
QY	181	gaatctattactttgtggaacagttgggtctcaaaattcatcagccacagttggaatg	240		
DB	246	GAATCTATTATTGTTGGAACAGTTGGCTCACAAAATTCATGCCCCACAGTTGGAATG	187		
QY	241	tctttattgttttctccaccggaggaacaccttaatgaacctgacagaagacagagaa	300		
DB	186	TCTTTATTGTTTTTCTCACCCGAGGAACAACCTTAATGAATGTCACAGAAGACAGAA	127		
QY	301	caaatcgtcgaaggcctagaagaactccagaagaagttcgcaggaggagacacttacatg	360		
DB	126	CRAATCGGTCAAGGCCTAGAGAAGTCCAGAAAGTTCTGCCAGGAGAGACACTTACATG	67		
QY	361	catgaagatttgaaggcgcagtgagcagatttattatgaacacagacagaggtacag	420		
DB	66	CATGAAGGATTTGAAGGGCCAGTGAGCAGATTTATTATGAAACACAGCAAGGCTACAGG	7		
QY	421	ac 422			
DB	6	ac 5			
RESULT	10				
BE146075					
LOCUS	BE146075	542 bp	mRNA	linear	EST 21-JUN-2000
DEFINITION	MR0-HT0209-280300-106-b03 HT0209				Homo sapiens cdna, mRNA sequence.
ACCESSION	BE146075				
VERSION	BE146075.1	GI:8608799			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
	1 (bases 1 to 542)				
	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,				
	Nagal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,				
	Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,				
	Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare				
	,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and				
	Simpson,A.J.				
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed				
	sequence tags				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	97 (7),	3491-3496	(2000)	
MEDLINE	20202663				
COMMENT	Contact: Simpson A.J.G.				
	Laboratory of Cancer Genetics				
	Ludwig Institute for Cancer Research				
	Rua Prof. Antonio Prudente 109, 4 andar,	01509-010,	Sao Paulo-SP,		
	Brazil				
	Tel: +55-11-2704922				
	Fax: +55-11-2707001				
	Email: asimpson@ludwig.org.br				
	This sequence was derived from the FAPESP/LICR Human Cancer Genome				
	Project. This entry can be seen in the following URL				
	( <a href="http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t3-MR0-HT0209-280300-106-b03&amp;t3=2000-03-28&amp;t4=1">http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t3-MR0-HT0209-280300-106-b03&amp;t3=2000-03-28&amp;t4=1</a> )				
	Seq primer: puc 18 forward				
	High quality sequence start: 51				
	High quality sequence stop: 131.				

FEATURES	source	Location/Qualifiers
	1. 542	
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone_lib="HT0209"
		/dev_stage="Adult"
		/note="Organ: head neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT	134 a 131 c 162 g 115 t	
ORIGIN		
Query Match	38.1%;	Score 420.4; DB 9; Length 542;
Best Local Similarity	99.8%;	Pred. No. 5e-104;
Matches 421;	Conservative	0; Mismatches 1; Indels 0; Gaps 0;
Qy 1	atggccacggcgagcgagagccctcgccatcggtccaggtccagtggtcttttggccact 60	
Db 105	ATGGCCACGGGAGCGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 164	
Qy 61	ctggtgctcatctgcgcgcgagcgcaaggcgagcgaggtgggggtccagcctgcttac 120	
Db 165	CTGGTGCTCATCTGCGCCGGCGAAGGGGAGCGCAGGAGGATGGGGGTCCAGCCTGCTAC 224	
Qy 121	ggcgatttgacctgtactctcatttggacaaatcaggaagtgtgctgcaccacttggaaat 180	
Db 225	GGCGATTGACCTCTACTTTCATTTTGGCAAAATCAGGAAGTGTCTGCACCACTTGGAAAT 284	
Qy 181	gaatctattacttgtggaacagttggtcacaattcatccagccacagttgagaatg 240	
Db 285	GAATCTATTACTTTGTGGAACAGTTGGCTCACAATTCATCGCCACACAGTTTGAATG 344	
Qy 241	tccttattgtttctccaccgcgaggaacaaccttaatgaaactgacagaagacagagaa 300	
Db 345	TCCTTTATGTGTTTCTCCACCGGAGAACAACTTAATGAAACTCACAGAAGACAGAGAA 404	
Qy 301	caaatccgtcaaggcctagaagaactccaagaagttctgcaggaggagagacattacatg 360	
Db 405	CAAAATCCGTCAAAGGCTTAGAAGAACTCCAGAAGTTCTGCCAGGAGGAGACACTTACATG 464	
Qy 361	catgaagatttgaaggccagtgagcagatttatgaaaaacagacaaggttacagg 420	
Db 465	CATGAAGATTGTAAGAGCCAGTGAGCAGATTATTATGAAACACAGACAAGGGTACCGG 524	
Qy 421	ac 422	
Db 525	AC 526	
RESULT 11		
BE145958/c		
LOCUS	BE145958	455 bp mRNA linear EST 21-JUN-2000
DEFINITION	MRO-HT0209-100100-103-f09 HT0209 Homo sapiens cDNA, mRNA sequence.	
ACCESSION	BE145958	
VERSION	BE145958.1	GI:8608682
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 465)	
	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva M. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.	
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed	



```

sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-MR0-HT0209-100
100-103-f09st3-2000-01-10st4=1)
Seq primer: puc 18 forward
High quality sequence start: 12
High quality sequence stop: 464.
Location/Qualifiers
FEATURES
    source
        1..465
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="HT0209"
            /dev_stage="Adult"
            /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
            Site_2: SmaI; A mini-library was made by cloning products
            derived from ORESTES PCR (U.S. Letters Patent application
            No. 196,716 - Ludwig Institute for Cancer Research)
            profiles into the pUC 18 vector. Reverse transcription of
            tissue mRNA and cDNA amplification were performed under
            low stringency conditions."
BASE COUNT 100 a 125 c 119 g 121 t
ORIGIN
Query Match 36.2%; Score 399.8; DB 9; Length 465;
Best Local Similarity 97.1%; Pred. No. 2e-98;
Matches 407; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 atggccacggcgagcgagagccctcgccatcggtctccagtgctctctttggccact 60
|||||
Db 431 ATGGCCACCGCGGACCGAGAGCCCTCGGCATCGGTTCCAGTGGCTCTCTTGGGCCACT 372
|||||

Qy 61 ctggtgctcatctgcgcgcaaggcgacgagggatgggggtccagcctgctac 120
|||||
Db 371 CTGTTGCTCATCTGCCCGCGGAGGGGTACGACGGAGATGGGGTCCAGCCTTGCTAC 312
|||||

Qy 121 ggcggattgacctgtacttctatttgacaaaatcaggaagtgtgctgcaccactggaat 180
|||||
Db 311 GCGGATTTTGACCTGTACTTCTATTGTCACAAATCAGGAAGTGTGCTGCACCACCTGGAAT 252
|||||

Qy 181 gaaatctattactttgttgagacagttggtccacaaatcaccagccacagttgagaatg 240
|||||
Db 251 GAAATCTATTACTTTGTGGAACAGTTGGCTCACAAATTCATCAGCCACAGTTGAGAATG 192
|||||

Qy 241 tctttattgtttctccaccggagaaacaccttaataatgaactgacagaagacagagaa 300
|||||
Db 191 TCCTTTATTGTTTCTCCACCGAGAACACCTTAATGAACCTGCACAGAGACAGAGAA 132
|||||

Qy 301 caaatcccgtaaggcctagagaactccagaaagtctgccaggaggagacacttacatg 360
|||||
Db 131 CAAATCCGTCAGGCGCTAGAGAACTCCAGAAAGTTCTGCCAGGAGGAGACACTTACATG 72
|||||

Qy 361 catgaagatttgaaagggccagtgagcagatttattatgaaaaacagacagaggtacag 419
|||||
Db 71 CATGAAGGATTTGAAAGGGCCAGTGAGCAGATTTTATTATGAAAACTAGACAAAGGTACAG 13
|||||

RESULT 12
BEI46355/c 487 bp mRNA linear EST 21-JUN-2000
LOCUS BEI46355
DEFINITION MR0-HT0209-010500-110-f09 HT0209 Homo sapiens cDNA, mRNA sequence.
ACCESSION BEI46355

```

```

VERSION BEI46355.1 GI:8609079
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 487)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-MR0-HT0209-010
500-110-f09st3-2000-05-01st4=1)
Seq primer: puc 18 forward
High quality sequence start: 30
High quality sequence stop: 487.
Location/Qualifiers
FEATURES
    source
        1..487
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="HT0209"
            /dev_stage="Adult"
            /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
            Site_2: SmaI; A mini-library was made by cloning products
            derived from ORESTES PCR (U.S. Letters Patent application
            No. 196,716 - Ludwig Institute for Cancer Research)
            profiles into the pUC 18 vector. Reverse transcription of
            tissue mRNA and cDNA amplification were performed under
            low stringency conditions."
BASE COUNT 104 a 147 c 121 g 115 t
ORIGIN
Query Match 34.9%; Score 384.8; DB 9; Length 487;
Best Local Similarity 99.2%; Pred. No. 2.6e-94;
Matches 397; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 1 atggccacggcgagcgagagccctcgccatcggtctccagtgctctctttggccact 60
|||||
Db 400 ATGGCCACCGCGGAGCGAGAGCCCTCGGCATCGGTTCCAGTGGCTCTCTTGGGCCACT 341
|||||

Qy 61 ctggtgctcatctgcgcgcaaggcgacgagggatgggggtccagcctgctac 120
|||||
Db 340 CTGTTGCTCATCTTGGCGCGGCAAGGGGACGCGAGGAGATGGGGTCCAGCCTTGCTAC 281
|||||

Qy 121 ggcggattgacctgtacttctatttgacaaaatcaggaagtgtgctgcaccactggaat 180
|||||
Db 280 GCGGATTTTGACCTGTACTTCTATTGTCACAAATCAGGAAGTGTGCTGCACCACCTGGAAT 221
|||||

Qy 181 gaaatctattactttgttgagacagttggtccacaaatcaccagccacagttgagaatg 240
|||||
Db 220 GAAATCTATTACTTTGTGGAACAGTTGGCTCACAAATTCATCAGCCACAGTTGAGAATG 161
|||||

Qy 241 tctttattgtttctccaccggagaaacaccttaataatgaactgacagaagacagagaa 300
|||||
Db 160 TCCTTTATTGTTTCTCCACCGAGGAACAGCCTTAATGAAACTGACAGAAGACAGAGAA 101
|||||

```



```

/clone_lib="NIH_MGC_14"
/tissue_type="renal cell adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: kidney; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
171 a 162 c 214 q 146 t

```

Query Match	34.1%	Score 376.4	DB 10	Length 693
Best Local Similarity	91.9%	Pred. No. 5.9e-92		
Matches 487	Conservative 0	Mismatches 31	Indels 12	Gaps 8
QY	18	gagagccctcggaatcgagctccagtgagctctcttttgccactctggtgctcatctgcgc	77	
DB	164	GAGCGCTCGGACATCGCTCCAGTGGCTCTCTTTGGCCACTCTGGTGCATCTGGCC	223	
QY	78	cgggcaaggggagcagcgagggagtggggggtccagcctgctacgcggaattgacctga	137	
DB	224	CGGCAAGGGGGAGCGAGGAGATGGGGTCCAGCCTGCTACGCGGATGGACCTGA	283	
QY	138	cttcattttggacaaatcaggaagtgtctgcacacctggaatgaaatctattacttgt	197	
DB	284	CTTCATTGGGACAAATCAGGAAGTGTCTGCACCACCTGGAATGAAATCTATTACTTGGT	343	
QY	198	ggaacagttaggtcacaaaattcatcagcccaagttgagaatgtcccttattgtttctc	257	
DB	344	GGAACAGTGGCTCACAAATTCATACGCCACAGTTGAGAAATGTCCTTATGGTTTCTC	403	
QY	258	cacccgaggaaacaccttaatgaactgacagaagacagagacaaatccctcaaggcct	317	
DB	404	CACCCGAGGAACACCTTATGAACCTGCAGAGACAGAGACAAATCCGTCAAGGCT	463	
QY	318	agaagaactccagaa-agttctgccagga-ggagacacttcacatgatgaa-ggatttga	374	
DB	464	AGAAGAACTCCAGAACAGTTCTGCAGGATGGAGACACTTACATGCATGAACGGATCGA	523	
QY	375	aagggccagtgagcagattttattgaaaaacagacagagggtacaggacagcagc--gtc	432	
DB	524	AAGGCCAGTGAGCAGATTTATTATGAACACAGACAAGGTTACAGGACAGCCGCGTCC	583	
QY	433	atcaattctttgactgatg----gagaactccatgaagatctcttttctattcca-gaga	487	
DB	584	ACCATTGGCTTGACCTTGATCGGAGAACTCCATGAAGATCTCTTAACTATTACAGAGA	643	
QY	488	ggggaggaat-aggtctcagatctt--ggtcgaattgtttactgtgttg	535	
DB	644	CGAGGCTAATAAGGCTTCGAAATCTTGGGTGCATCTGGTTAACTGGGGTGG	693	

RESULT	15
AI084806/c	
LOCUS	. linear EST 01-OCT-1998
DEFINITION	oz80d12.xl Soares_senescent.fibroblasts_NBHSF Homo sapiens cDNA clone IMAGE:166165 3', mRNA sequence.
ACCESSION	AI084806
VERSION	AI084806
KEYWORDS	GI:3423229
SOURCE	EST.
ORGANISM	human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 554)
AUTHORS	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)

```

COMMENT
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 721 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 455.
Location/Qualifiers
1. .554
    . /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_lib="IMAGE:1681655"
      /tissue_type="senescent fibroblast"
      /lab_host="DH10B (ampicillin resistant)"
      /note="Vector: pT73D (Pharmacia) with a modified
polylinker V_TYPE: phagemid; Site_1: Not I - oligo(dT)
; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5',
TGTTACCATCTCAAGTGGCGCGCGCATTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo."
152 a 110 c 125 g 167 t
BASE COUNT
ORIGIN

```

	Query Match	33.88;	Score 373.6;	DB 9;	Length 554;
	Best Local Similarity	98.99;	Pred. No. 3.2e-91;		
	Matches 376;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;
Qy	578	acagtaaggaacatggtgttcccgtagaatgaegcgcttccagggctctgcaagcgcatcc	637		
Db	554	ACAGTAGGATCATGTGTTCCCGTGAATGACGGCTTTCAGGCTCTCGAAGCATCATCC	495		
Qy	638	actcaattttgaagaagtcctgtgcatacgaaattctagcagctggaaccatccaccatatgtg	697		
Db	494	ACTCAATTTTGAAGAAGTCCCTGCATCGCAATTCCTAGCAGCTGAACCATCCACCATATGTG	435		
Qy	698	cagagagatcattcaagttgctgagagaaagcgcttccgacatgcccgcaactgg	757		
Db	434	CAGGAGAGTCAATTCAGTTGCTGTGAGAGAAACGGCTCCGCACATGCCCGCAACGTGG	375		
Qy	758	acagggtcctctcgagcttcaagatcaatgactcgggtcacactcaatgagaagccctttt	817		
Db	374	ACAGGGTCTCTCGCAGCTTCAAGATCAATGACTCGGTGCACACTCAATGAGAAGCCCTTTT	315		
Qy	818	ctgtggaagacacattattactgtgtccagcgctactcttaagaagagtgccatgaag	877		
Db	314	CTGTGGAAGATACATTTATTTACTGTGTCCAGCGCCTATCTTAAAGAAGTTGGCATGAAG	255		
Qy	878	ctgcactccaggtcagcatgacgatgacgtcctctcttttatctccagttctctgcatcatca	937		
Db	254	CTGCACCTCCAGGTGAGCATGAACGATGSCCTCTCTTTTATCTCCAGTTCTGTCTCATCATCA	195		
Qy	938	ccacacacacatgttctgac	957		
Db	194	CCACCACACACTGTGATGGC	175		

Search completed: August 8, 2002, 23:02:59  
Job time: 10146 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 9, 2002, 00:30:14 ; Search time 413.49 seconds  
(without alignments)  
4584.086 Million cell updates/sec

Title: US-09-970-076-1-copy\_104\_1207  
Perfect score: 1104  
Sequence: 1 atggccacgcggagcgag.....gtgaggaaataataaaaa 1104

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 1736436 seqs, 858457221 residues  
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :			
N_Geneseq_032802.*			
1:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*		
2:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*		
3:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*		
4:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*		
5:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*		
6:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*		
7:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*		
8:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*		
9:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*		
10:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*		
11:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*		
12:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*		
13:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*		
14:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*		
15:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*		
16:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*		
17:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*		
18:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*		
19:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*		
20:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*		
21:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*		
22:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*		
23:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*		
24:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1090.6	98.8	2447	22	AA05303 Human secreted pro
2	1089.4	98.7	2086	22	AA05334 Human secreted pro
3	949.4	86.0	2272	21	AAA47455 Human TANGO 197 co
4	803.8	72.8	1402	22	AAI59918 Human polynucleoti
5	803.8	72.8	1609	22	AAI58132 Human polynucleoti
6	787	71.3	1436	22	AAH14331 Human cDNA sequenc
7	453.2	41.1	4417	21	AAA47479 Murine TANGO 197 c
8	349	31.6	356	22	AAI19043 Human breast cance
9	348	31.5	355	22	AAI08133 Human breast cance

10	346	31.3	346	22	AAI25878 Human breast cance
11	344.6	31.2	3677	21	AAA75149 cDNA encoding a hu
12	343	31.1	3677	21	AAA75157 cDNA clone encodin
13	343	31.1	3677	21	AAA75158 cDNA clone encodin
14	343	31.1	3677	21	AAA75159 cDNA clone encodin
15	341.8	31.0	3501	21	AAA75162 cDNA clone encodin
16	341.2	30.9	2239	22	AAS31233 Human cDNA encodin
17	340.2	30.8	3501	21	AAA75150 cDNA encoding a mu
18	340.2	30.8	3501	21	AAA75160 cDNA clone encodin
19	338.6	30.7	3501	21	AAA75161 cDNA clone encodin
20	277.6	25.1	464	22	AAH06188 Human cDNA clone (
21	186.4	16.9	211	20	AAH86628 EST clone G8874.
22	136.8	12.4	350	21	AAZ94448 Cartilage-associat
23	113	10.2	811	22	AAH08055 Human cDNA clone (
24	113	10.2	1189	22	AAH15137 Human cDNA sequenc
25	96.8	8.8	390	22	AAK57088 Human immune/haema
26	59.6	5.4	188	21	AAA45398 Human secreted exp
27	41.4	3.8	270	22	AAI00358 Human reproductive
28	37.4	3.4	3489	21	AAA30290 Kaposi's sarcoma-a
29	37.4	3.4	3489	22	AAF82901 Nucleotide sequenc
30	37.4	3.4	32207	20	AAV73805 KSHV LTR DNA (nucl
31	37.4	3.4	32217	22	AA541738 Genomic sequence #
32	37.4	3.4	137507	19	AAV19941 KSHV long unique c
33	37.2	3.4	488	21	AAA16363 Human colon cancer
34	36.6	3.3	2533	23	AA594411 DNA encoding novel
35	36.6	3.3	2833	21	AA58396 Human PRO2198 nucl
36	36.6	3.3	4969	23	ABL20345 Drosophila melanog
37	36.6	3.3	5553	23	ABLI0259 Drosophila melanog
38	36.6	3.3	6741	21	AAAI0595 Gene encoding a su
39	36.6	3.3	14487	23	ABLI0258 Drosophila melanog
40	36.6	3.3	14768	23	ABLI0258 Drosophila melanog
41	36.4	3.3	783	21	AACT1316 Single nucleotide
42	36.4	3.3	1113	24	AA516932 Human epithelin-like
43	36.4	3.3	5250	22	AAK52371 Human polynucleoti
44	36.4	3.3	5254	22	AAK53355 Human polynucleoti
45	36.2	3.3	2198	21	AAF08319 Fusarium venenatum

ALIGNMENTS

RESULT 1  
AA05303  
ID AAD05303 standard; cDNA; 2447 BP.

AC AAD05303;

XX 17-JUL-2001 (first entry)

DT Human secreted protein-encoding gene 4 cDNA clone HMLFR02, SEQ ID NO:14.

DE Human; secreted protein; proliferative disorder; cancer; tumour;  
KW fetal abnormality; developmental abnormality; haematopoietic disorder;  
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
KW gastrointestinal disorder; pregnancy-related disorder;  
KW endocrine disorder; infection; wound healing; vulnerability;  
KW cell culture; chemotaxis; food additive; gene therapy;  
KW binding partner identification; chromosome 19; ss.

OS Homo sapiens.

XX	Key	Location/Qualifiers
FT	CDS	140..1351
FT		/*tag= a
FT		/product= "Human secreted protein"
FT	sig_peptide	140..220
FT		/*tag= b
FT	mat_peptide	221..1348
FT		/*tag= c

/product="Mature human secreted protein"

	/product= "Mature human secreted protein"
XX	WO200134626-A1.
XX	17-MAY-2001.
XX	01-NOV-2000; 2000WO-US30045.
XX	05-NOV-1999; 99US-0163581.
XX	30-JUN-2000; 2000US-0215133.
XX	(HUMA-) HUMAN GENOME SCI INC.
XX	Ruben SM, Komatsoulis GA, Moore PA, Birse CE, Ni J;
XX	WPI; 2001-308778/32.
XX	P-PSDB; AAE01439.
XX	New nucleic acid molecules encoding 28 human secreted proteins for
XX	diagnosing, preventing, treating or ameliorating medical conditions and
XX	used as food additives or preservatives -
XX	Claim 1; Page 425-426; 562pp; English.
XX	AAD05300-AA0D05379 represent cDNAs corresponding to 28 human secreted
XX	protein genes, and AAE01436-AAE01513 represent the proteins they encode
XX	AAE01514-AAE01544 represent human secreted protein fragments or variants
XX	The genes and their secreted proteins are useful for preventing,
XX	treatment or ameliorating medical conditions, e.g., by protein or gene
XX	therapy. Pathological conditions can be diagnosed by determining the
XX	amount of the new protein in a sample or by determining the presence of
XX	mutations in the new genes. Specific uses are described for each of the
XX	28 genes, based on the tissues in which they are most highly expressed,
XX	and include developing products for the diagnosis or treatment of
XX	proliferative disorders, cancer, tumours, foetal and developmental
XX	abnormalities, haematopoietic disorders, diseases of the immune system,
XX	AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
XX	allergies, neurological disorders (e.g., Alzheimer's disease,
XX	Parkinson's disease), cognitive disorders, schizophrenia, asthma,
XX	skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
XX	cardiovascular disorders, angioneurotic disorders, kidney disorders,
XX	gastrointestinal disorders, pregnancy-related disorders, endocrine
XX	disorders, and infections. The proteins can also be used to aid wound
XX	healing and epithelial cell proliferation, to prevent skin aging due to
XX	sunburn, to maintain organs before transplantation, for supporting cell
XX	culture of primary tissues, to regenerate tissues, to identify their
XX	cognate ligands or binding partners, and in chemotaxis, and can be used
XX	as a food additive or preservative to modify storage properties.
XX	Antibodies specific for a protein of the invention can be used in
XX	alleviating symptoms associated with the disorders mentioned above, and
XX	in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
XX	immunosorbent assay (ELISA). The present sequence represents a human
XX	secreted protein-encoding cDNA of the invention.
XX	Sequence 2447 BP; 673 A; 562 C; 587 G; 625 T; 0 other;
SO	

QY	181	gaaatctattactttgtggaacagttgcttcacaaattcatcagccacacagttgagaatg	240
Db	320	gaaatctattactttgtggaacagttgcttcacaaattcatcagccacacagttgagaatg	379
QY	241	tcctttattgttttctccaccgaggaacaaccttaataaactgcagagaacagagaa	300
Db	380	tcctttattgttttctccaccgaggaacaaccttaataaactgcagagaacagagaa	439
QY	301	caaatccgctcaagcgcttagaagaactccagaaaaagtctgcaggagagacacttcatg	360
Db	440	caaatccgctcaagcgcttagaagaactccagaaaaagtctgcaggagagacacttcatg	499
QY	361	catgaagagattgaaaggccagtgagcagatttattatgaaaacagacaagggtacagg	420
Db	500	catgaagagattgaaaggccagtgagcagatttattatgaaaacagacaagggtacagg	559
QY	421	acagccagcgctcatcattgctttgactgatggagaactccatgaagatctctttttctat	480
Db	560	acagccagcgctcatcattgctttgactgatggagaactccatgaagatctctttttctat	619
QY	481	tcagaagaggaggtcaatagatctcgagatcttggtgcaattgtttactgtgttggtgtg	540
Db	620	tcagaagaggaggtcaatagatctcgagatcttggtgcaattgtttactgtgttggtgtg	679
QY	541	aaagatttcaatgagacacagctggcccggtattgcggaagtaagatcatgtgtttccc	600
Db	680	aaagatttcaatgagacacagctggcccggtattgcggaagtaagatcatgtgtttccc	739
QY	601	gtgaatgaagcggtttcaagctctgaaggcatccactcaattttgaagaagtcctgc	660
Db	740	gtgaatgaagcggtttcaagctctgaaggcatccactcaattttgaagaagtcctgc	799
QY	661	atcgaaattctgacagctgaaccatccaccatattgtaggagagtcatttcaagttgtc	720
Db	800	atcgaaattctgacagctgaaccatccaccatattgtaggagagtcatttcaagttgtc	859
QY	721	gtgagaggaaaacggcttccgacatgcccgcaacgctggacaggtctctgcagcttcaag	780
Db	860	gtgagaggaaaacggcttccgacatgcccgcaacgctggacaggtctctgcagcttcaag	919
QY	781	atcaatgaactcggctcacactcaatgagaagccctttctgtggaagacacttattactg	840
Db	920	atcaatgaactcggctcacactcaatgagaagccctttctgtggaagacacttattactg	979
QY	841	tgtccagcgctatctttaaaagaagtggcagtgaagaagctgcactccaggtcagcatgaac	900
Db	980	tgtccagcgctatctttaaaagaagtggcagtgaagaagctgcactccaggtcagcatgaac	1039
QY	901	gatggcctctttttatctccagttctgctcatcatcaccaccacacactgtttcgacggt	960
Db	1040	gatggcctctttttatctccagttctgctcatcatcaccaccacacactgtttcgacggt	1099
QY	961	tcacatctggccatcgccctgctgactgttctctgctcctagccctggtctcctctg	1020
Db	1100	tcacatctggccatcgccctgctgactgttctctgctcctagccctggtctcctctg	1159
QY	1021	tggtttcggccctctgctgcactgtgattatcaaggaggtccctcaccacctgcccag	1080
Db	1160	tggtttcggccctctgctgcactgtgattatcaaggaggtccctcaccacctgcccag	1219
QY	1081	gaagagtgaagaaataa	1097
Db	1220	gaagagtgaagtaagtga	1236
RESULT 2			
AAD05334			
ID	AAD05334 standard; cDNA; 2086 bp.		
XX	AC		
XX	AAD05334;		
DT	17-JUL-2001 (first entry)		
XX	XX		

DE Human secreted protein-encoding gene 4 CDNA clone HWLFR02, SEQ ID NO:45.

XX Human; secreted protein; proliferative disorder; cancer; tumour;

KW foetal abnormality; developmental abnormality; haematopoietic disorder;

KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;

KW inflammation; allergy; neurological disorder; Alzheimer's disease;

KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;

KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;

KW cardiovascular disorder; angiogenic disorder; kidney disorder;

KW gastrointestinal disorder; pregnancy-related disorder;

KW endocrine disorder; infection; wound healing; vulnerability;

KW cell culture; chemotaxis; food additive; gene therapy;

KW binding partner identification; chromosome 19; ss.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 166..1377

FT /tag= a

FT /product= "Human secreted protein"

FT /transl\_except= (pos:688..690, aa:Xaa)

FT /transl\_except= (pos:1123..1125, aa:Xaa)

FT /transl\_except= (pos:1156..1158, aa:Xaa)

FT /transl\_except= (pos:1267..1269, aa:Xaa)

FT /note= "Xaa equals any of the twenty naturally occurring L-amino acids"

FT sig\_peptide 166..246

FT /tag= b

FT mat\_peptide 247..1374

FT /tag= c

FT /product= "Mature human secreted protein"

XX

PN WO200134626-A1.

XX

PD 17-MAY-2001.

XX

PF 01-NOV-2000; 2000WO-US30045.

PR 05-NOV-1999; 99US-0163581.

PR 30-JUN-2000; 2000US-0215133.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Ruben SM, Komatsoulis GA, Moore PA, Birse CE, Ni J;

XX

DR WPI: 2001-308778/32.

DR P-PSDB; AA001469.

XX

PT New nucleic acid molecules encoding 28 human secreted proteins for

PT diagnosing, preventing, treating or ameliorating medical conditions and

PT used as food additives or preservatives -

XX

PS Claim 1; Page 449; 562pp; English.

XX

CC AAD05300-AA005379 represent cDNAs corresponding to 28 human secreted

CC protein genes, and AA001436-AA001513 represent the proteins they encode.

CC AA001514-AA001544 represent human secreted protein fragments or variants.

CC The genes and their secreted proteins are useful for preventing,

CC treating or ameliorating medical conditions, e.g., by protein or gene

CC therapy. Pathological conditions can be diagnosed by determining the

CC amount of the new protein in a sample or by determining the presence of

CC mutations in the new genes. Specific uses are described for each of the

CC 28 genes, based on the tissues in which they are most highly expressed,

CC and include developing products for the diagnosis or treatment of

CC proliferative disorders, cancer, tumours, foetal and developmental

CC abnormalities, haematopoietic disorders, diseases of the immune system,

CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,

CC allergies, neurological disorders (e.g., Alzheimer's disease,

CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,

CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,

CC cardiovascular disorders, angiogenic disorders, kidney disorders,

CC gastrointestinal disorders, pregnancy-related disorders, endocrine

CC disorders, and infections. The proteins can also be used to aid wound

CC healing and epithelial cell proliferation, to prevent skin aging due to

CC sunburn, to maintain organs before transplantation, for supporting cell

CC culture of primary tissues, to regenerate tissues, to identify their

CC cognate ligands or binding partners, and in chemotaxis, and can be used

CC as a food additive or preservative to modify storage properties.

CC Antibodies specific for a protein of the invention can be used in

CC alleviating symptoms associated with the disorders mentioned above, and

CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked

CC immunosorbent assay (ELISA). The present sequence represents a human

CC secreted protein-encoding cDNA of the invention.

XX

SQ Sequence 2086 BP; 544 A; 491 C; 538 G; 501 T; 12 other;

Query Match 98.7%; Score 1089.4; DB 22; Length 2086;

Best Local Similarity 99.4%; Pred. No. 0;

Matches 1090; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 atggccacgagcgagcgagagccctcgccatcggtctccagtggtctctttggccact 60

DB 166 atggccacgagcgagcgagagccctcgccatcggtctccagtggtctctttggccact 225

QY 61 ctggtgctcatctgcgccgggcaaggggacgagggagatgggggtccagcctctctac 120

DB 226 ctggtgctcatctgcgccgggcaaggggacgagggagatgggggtccagcctctctac 285

QY 121 ggcgatttgacgtgacttcttcttggacaaatcaggaagtgctgcaccactggaat 180

DB 286 ggcgatttgacgtgacttcttcttggacaaatcaggaagtgctgcaccactggaat 345

QY 181 gaattctattcttggagacagtggtgcacaaatctcatgcaccacagtgaggaatg 240

DB 346 gaattctattcttggagacagtggtgcacaaatctcatgcaccacagtgaggaatg 405

QY 241 tcctttattgtttctccaccggaggaacaccttaataagaaacagagacagagaa 300

DB 406 tcctttattgtttctccaccggaggaacaccttaataagaaacagagacagagaa 465

QY 301 caatccgtcaagccttagaagaactccagaagttctccagagagagacacttacatg 360

DB 466 caatccgtcaagccttagaagaactccagaagttctccagagagagacacttacatg 525

QY 361 catgaaggatttgaaggccagtgagcagatttattatgaacacagacagaggtacagg 420

DB 526 catgaaggatttgaaggccagtgagcagatttattatgaacacagacagaggtacagg 585

QY 421 acagccagcgtcatctgtttgactgaggaacatccatgaagatctcttttttat 480

DB 586 acagccagcgtcatctgtttgactgaggaacatccatgaagatctcttttttat 645

QY 481 tcagagaggagagcctaagtgctcagagatcttggtgcaattgtttactgtgtgtgtg 540

DB 646 tcagagaggagagcctaagtgctcagagatcttggtgcaattgtttactgtgtgtgtg 705

QY 541 aaagatttcaatgagacacagcgtgcccggattgctgagacagaggaagatcatgtgttccc 600

DB 706 aaagatttcaatgagacacagcgtgcccggattgctgagacagaggaagatcatgtgttccc 765

QY 601 gtaagtacggcttccaggtctgcgaaggcatcatccactcaattttgagaagaagctctgc 660

DB 766 gtaagtacggcttccaggtctgcgaaggcatcatccactcaattttgagaagaagctctgc 825

QY 661 atcgaaattctagcagctgaaccatccaccatattgtcagagagagtcatttcaagttgtc 720

DB 826 atcgaaattctagcagctgaaccatccaccatattgtcagagagagtcatttcaagttgtc 885

QY 721 gtgagaggaacggcttccgacatgcccgaagtgagacaggggtccctctgacagttcaag 780

DB 886 gtgagaggaacggcttccgacatgcccgaagtgagacaggggtccctctgacagttcaag 945

QY 781 atcaatgactcgttcacactcaatgagaagccctttctgtggaagacacttattactg 840

DB 946 atcaatgactcgttcacactcaatgagaagccctttctgtggaagacacttattactg 1005

QY 841 tgtcagcgccctattttaaaagaattggcatgaagtgcaactccaggtcagcatgaac 900  
Db 1006 tgtcagcgccctattttaaaagaattggcatgaagtgcaactccaggtcagcatgaac 1065  
QY 901 gatggcctcttttatctccagttctgtcatcatcacacacacacacactgttctgacgtg 960  
Db 1066 gatggcctcttttatctccagttctgtcatcatcacacacacacacactgttctgacgtg 1125  
QY 961 tcatctctggccatcgccctgtgatccctgttctctgtctctagccctggctctctctg 1020  
Db 1126 tcatctctggccatcgccctgtgatccctgttctctgtctctagccctggctctctctg 1185  
QY 1021 tggttctggccctctgtctgacgtgtgattatcaaggaggtccctccacccctgccag 1080  
Db 1186 tggttctggccctctgtctgacgtgtgattatcaaggaggtccctccacccctgccag 1245  
QY 1081 gagagtgaggaataa 1097  
Db 1246 gagagtgaggaagtga 1262  
RESULT 3  
AAA47455  
ID AAA47455 standard; cDNA; 2272 BP.  
XX  
AC AAA47455;  
XX  
DT 20-OCT-2000 (first entry)  
XX  
DE Human TANGO 197 coding sequence.  
XX  
KW TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;  
KW graft versus-host diseases; rheumatoid arthritis; psoriasis;  
KW inflammatory bowel disease; septic shock; ulcerative colitis;  
KW Crohn's disease; chronic myelogenous leukemia; cancer; liver  
KW disease; Hodgkin's disease; osteoarthritis; Lyme's disease;  
KW cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;  
KW systemic lupus erythematosus; transgenic animal; diagnosis;  
KW prognosis; prophylactic; therapeutic; human; ds.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 213..1214  
FT /\*tag= a  
FT /product= TANGO 197  
XX  
PN WO200039284-A1.  
XX  
XX 06-JUL-2000.  
XX  
XX 23-DEC-1999; 99WO-US31025.  
XX  
XX 30-DEC-1998; 98US-0223546.  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
XX Holtzman DA;  
XX  
XX WPI; 2000-465743/40.  
XX P-PSDB; AAB01422.  
XX  
XX Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,  
XX 224 and 239 polypeptides useful for the treatment of asthma, rheumatoid  
XX arthritis, psoriasis and autoimmune diseases  
XX  
XX Claim 1; Fig 4; 209pp; English.  
XX  
XX Nucleic acids encoding TANGO polypeptides are useful as modulating  
XX agents for regulating cellular processes like asthma, graft  
XX versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory  
XX bowel disease, septic shock, ulcerative colitis, Crohn's disease,

CC chronic myelogenous leukemia, cancer, liver disease, Hodgkin's  
CC disease, osteoarthritis, Lyme's disease, cachexia and autoimmune  
CC diseases e.g. myasthenia gravis, autoimmune diabetes and systemic  
CC lupus erythematosus. The nucleic acids are also useful for producing  
CC transgenic animals and the TANGO polypeptides themselves. Partial  
CC TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in  
CC pharmacogenomics and for diagnostic assays, prognostic assays,  
CC forensic biology, for monitoring clinical trials. TANGO  
CC polypeptides are suitable for both prophylactic and therapeutic  
CC methods for treating a subject at risk of a disorder or having a  
CC disorder associated with aberrant TANGO expression. A wide range  
CC of cellular disorders can be treated.  
XX  
SQ Sequence 2272 BP; 638 A; 554 C; 558 G; 522 T; 0 other;  
  
Query Match 86.0%; Score 949.4; DB 21; Length 2272;  
Best Local Similarity 99.9%; Pred. NO. 2.5e-280;  
Matches 950; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 atggccacggcgagcgagagccctcgcatcggcttcacgtggctctctttggccact 60  
Db 213 atggccacggcgagcgagagccctcgcatcggcttcacgtggctctctttggccact 272  
QY 61 ctggtgtctcatctgcgcgggcaagggggacgcaggaggatggggggtccagcctgctac 120  
Db 273 ctggtgtctcatctgcgcgggcaagggggacgcaggaggatggggggtccagcctgctac 332  
QY 121 ggcggatttgacctgaacttcattttggacaactcaggaagtgtctgcaccactggaat 180  
Db 333 ggcggatttgacctgaacttcattttggacaactcaggaagtgtctgcaccactggaat 392  
QY 181 gaaatctattctgttggaacagttggctcacaattcatcagccacagttggaatg 240  
Db 393 gaaatctattctgttggaacagttggctcacaattcatcagccacagttggaatg 452  
QY 241 tctttattgtttctccaccgaggaacacacttaataaactgaacagacagacagaa 300  
Db 453 tctttattgtttctccaccgaggaacacacttaataaactgaacagacagacagaa 512  
QY 301 caaatccgtcaaggcctagagaactccagaagtcttccaggaggagacacttacatg 360  
Db 513 caaatccgtcaaggcctagagaactccagaagtcttccaggaggagacacttacatg 572  
QY 361 catgaaggatttgaaaggccagtgagcgatttattatgaaaacagacaaaggtcacag 420  
Db 573 catgaaggatttgaaaggccagtgagcgatttattatgaaaacagacaaaggtcacag 632  
QY 421 acagccagcgctcatcttcttgactgatggagaactccatgaagatctctttctat 480  
Db 633 acagccagcgctcatcttcttgactgatggagaactccatgaagatctctttctat 692  
QY 481 tcagagaggaggctaataggtctcgagatcttgggtgcaattgttactgtgtgtgtg 540  
Db 693 tcagagaggaggctaataggtctcgagatcttgggtgcaattgttactgtgtgtgtg 752  
QY 541 aaagatttcaatgagacacagctggcccggttgcggacagtagtaagatcatgtgtttccc 600  
Db 753 aaagatttcaatgagacacagctggcccggttgcggacagtagtaagatcatgtgtttccc 812  
QY 601 gtgaatgacggcttccagctctgcaggcatcatccactcaatttttgaagaagtcctgc 660  
Db 813 gtgaatgacggcttccagctctgcaggcatcatccactcaatttttgaagaagtcctgc 872  
QY 661 atcgaaattctagcagtgaacctccaccatgatgtgcaggagagtagtoatttcaagttgtc 720  
Db 873 atcgaaattctagcagtgaacctccaccatgatgtgcaggagagtagtoatttcaagttgtc 932  
QY 721 gtgagaggaacggcttcgacatgcccgaactggagacaggtctctctcagcttcaag 780  
Db 933 gtgagaggaacggcttcgacatgcccgaactggagacaggtctctctcagcttcaag 992  
QY 781 atcaatgactcgttccacactcaatgagaagcccttttctgtggaagacacttattactg 840



Db 993 atcaatgactcgtgcacatcaatgagaagccctttctgtggaagatactatttactg 1052  
Qy 841 tctccagcgcctatcttaaaagaagtggcatgaaagctgcactccaggtcagcatgaac 900  
Db 1053 tctccagcgcctatcttaaaagaagtggcatgaaagctgcactccaggtcagcatgaac 1112  
Qy 901 gatggcctctctttatctccagttctgtcatcatcacacacacactgt 951  
Db 1113 gatggcctctctttatctccagttctgtcatcatcacacacacactgt 1163  
RESULT 4  
ID AAI59918  
AAI59918 standard; cDNA; 1402 BP.  
XX  
AC AAI59918;  
XX  
DT 22-OCT-2001 (first entry)  
XX Human polynucleotide SEQ ID NO 3907.  
XX  
DE Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; SS.  
XX  
OS Homo sapiens.  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US34263.  
XX  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX  
XX WPI: 2001-442253/47.  
DR P-PSDB; AAM40762.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
PT  
XX  
PS Claim 1; SEQ ID NO 3907; 10078pp; English.  
XX  
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and

CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 1402 BP; 360 A; 354 C; 399 G; 289 T; 0 other;  
Query Match 72.8%; Score 803.8; DB 22; Length 1402;  
Best Local Similarity 99.8%; Pred No. 8.8e-236;  
Matches 805; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 atggccacggcgagcgagagccctcggcatcggtctccagtgcctctcttggccact 60  
Db 309 atggccacggcgagcgagagccctcggcatcggtctccagtgcctctcttggccact 368  
Qy 61 ctggtgctcatctgcgcggcgcaagggggacgcagggagagatgggggtccacactgctac 120  
Db 369 ctggtgctcatctgcgcggcgcaagggggacgcagggagatgggggtccacactgctac 428  
Qy 121 ggcgatttgacctgacttctcatttggacaatcagggaagtgtgctgaccactggaat 180  
Db 429 ggcgatttgacctgacttctcatttggacaatcagggaagtgtgctgaccactggaat 488  
Qy 181 gaaatctattactttgtggaacagttggctcacaattcatcagccacagttgagaatg 240  
Db 489 gaaatctattactttgtggaacagttggctcacaattcatcagccacagttgagaatg 548  
Qy 241 tctttattgtttctccaccgaggaacacacttaataagaactgacagaagacagagaa 300  
Db 549 tctttattgtttctccaccgaggaacacacttaataagaactgacagaagacagagaa 608  
Qy 301 caaatccgtcaaggcctagaagaactccagaagaagtctccaggaggagacacttacaag 360  
Db 609 caaatccgtcaaggcctagaagaactccagaagaagtctccaggaggagacacttacaag 668  
Qy 361 catgaaggatttgaaagggccagtgagcagattttattatgaataacagacaagggtacagg 420  
Db 669 catgaaggatttgaaagggccagtgagcagattttattatgaataacagacaagggtacagg 728  
Qy 421 acagccagcgtcatcatctgtttgactgatggagaactccatgaagatctcttttctat 480  
Db 729 acagccagcgtcatcatctgtttgactgatggagaactccatgaagatctcttttctat 788  
Qy 481 tcagagagggaggctaaataggtctctgagatctgtgcaattgtttactgtgtgtgtgtg 540  
Db 789 tcagagagggaggctaaataggtctctgagatctgtgcaattgtttactgtgtgtgtgtg 848  
Qy 541 aaagatttcaatgagacacagctggcccggttggtgggacagtaaggatcatgtgtttccc 600  
Db 849 aaagatttcaatgagacacagctggcccggttggtgggacagtaaggatcatgtgtttccc 908  
Qy 601 gtgaatgacggcttttcaggctctgcaaggcatccatccactcaattttgaaagactctgc 660  
Db 909 gtgaatgacggcttttcaggctctgcaaggcatccatccactcaattttgaaagactctgc 968  
Qy 661 atcgaaattctagcagctgaaccatcccatatgtgagagagagtcatttcaagtgttc 720  
Db 969 atcgaaattctagcagctgaaccatcccatatgtgagagagagtcatttcaagtgttc 1028  
Qy 721 gtgagaggaacacggtctccgacatgcccgcaacgctggacaggggtctctcgcagcttcaag 780  
Db 1029 gtgagaggaacacggtctccgacatgcccgcaacgctggacaggggtctctcgcagcttcaag 1088  
Qy 781 atcaatgactcgtgcacactcaatgag 807  
Db 1089 atcaatgactcgtgcacactcagtaag 1115  
RESULT 5  
AAI58132  
ID AAI58132 standard; cDNA; 1609 BP.  
XX  
AC AAI58132;



PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI: 2001-318749/34.  
 XX  
 PR Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs .  
 XX  
 PS Claim 8; SEQ ID 11705; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 1436 BP; 375 A; 349 C; 401 G; 311 T; 0 other;

Db 503 catgaagattgaaagggccagtgagcagatttattatgaaacagacaagggtacagg 562  
 Qy 421 acagccagcgcacatcattgcttgcactgatggagaacctccatgaagatctctttctat 480  
 Db 563 acagctagcgtcatcattgcttgcactgatggagaacctccatgaagatctctttctat 622  
 Qy 481 tcagagagggaggctcaataggtctcgcagatcttgggtgcaattgttactgtgtgtgtg 540  
 Db 623 tcagagagggaggctcaataggtctcgcagatcttgggtgcaattgttactgtgtgtgtg 682  
 Qy 541 aaagatttcaatgagacacagctggcccgagttgaggagtaagataatgtgtttccc 600  
 Db 683 aaagatttcaatgagacacagctggcccgagttgaggagtaagataatgtgtttccc 742  
 Qy 601 gtgaatgagcggcttccagcgtctcgaaggcatcaccactcaattttgaaagaagtcctgc 660  
 Db 743 gtgaatgagcggcttccagcgtctcgaaggcatcaccactcaattttgaaagaagtcctgc 802  
 Qy 661 atcgaattcttagcagctgaaccatccaccatattgagagagatcatttcaagtgttc 720  
 Db 803 atcgaattcttagcagctgaaccatccaccatattgagagagatcatttcaagtgttc 862  
 Qy 721 gtgagagaaacggcttccgcacatgcccgcaacgtggacaggggtcctctgcagctcaag 780  
 Db 863 gtgagagaaacggcttccgcacatgcccgcaacgtggacaggggtcctctgcagctcaag 922  
 Qy 781 atcaatgactcgtgcacactcaatgag 807  
 Db 923 atcaatgactcgtgcacactcaatgag 949

RESULT 7  
 AAA47479  
 ID AAA47479 standard; cDNA: 4417 BP.  
 XX  
 AC AAA47479;  
 XX  
 DT 20-OCT-2000 (first entry)  
 XX  
 DE Murine TANGO 197 coding sequence.  
 XX  
 KW TANGO; 128; 140; 197; 212; 213; 234; 239; modulating agent; asthma;  
 KW graft versus-host diseases; rheumatoid arthritis; psoriasis;  
 KW inflammatory bowel disease; septic shock; ulcerative colitis;  
 KW Crohn's disease; chronic myelogenous leukemia; cancer; liver  
 KW disease; Hodgkin's disease; osteoarthritis; Lyme's disease;  
 KW cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;  
 KW systemic lupus erythematosus; transgenic animal; diagnosis;  
 KW prognosis; prophylactic; therapeutic; mouse; ds.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 CDS 3..1148  
 FT /\*tag= a  
 FT /product= TANGO 197  
 XX  
 PN WO200039284-A1.  
 XX  
 PD 06-JUL-2000.  
 XX  
 PF 23-DEC-1999; 99WO-US31025.  
 XX  
 PR 30-DEC-1998; 98US-0223546.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Holtzman DA;  
 XX  
 PI WPI: 2000-465743/40.  
 DR P-PSDB; AAB01428.  
 XX

PT Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,  
PT 224 and 239 polypeptides useful for the treatment of asthma, rheumatoid  
PT arthritis, psoriasis and autoimmune diseases  
XX  
XX  
XX Claim 1; Fig 27; 209pp; English.  
XX  
XX Nucleic acids encoding TANGO polypeptides are useful as modulating  
CC agents for regulating cellular processes like asthma, graft  
CC versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory  
CC bowel disease, septic shock, ulcerative colitis, Crohn's disease,  
CC chronic myelogenous leukemia, cancer, liver disease, Hodgkin's  
CC disease, osteoarthritis, Lyme's disease, cachexia and autoimmune  
CC diseases e.g. myasthenia gravis, autoimmune diabetes and systemic  
CC lupus erythematosus. The nucleic acids are also useful for producing  
CC transgenic animals and the TANGO polypeptides themselves. Partial  
CC TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in  
CC forensic biology; for diagnostic assays, prognostic assays,  
CC pharmacogenomics and for monitoring clinical trials. TANGO  
CC polypeptides are suitable for both prophylactic and therapeutic  
CC methods for treating a subject at risk of a disorder or having a  
CC disorder associated with aberrant TANGO expression. A wide range  
CC of cellular disorders can be treated.  
XX  
XX SQ Sequence 4417 BP; 1214 A; 1115 C; 994 G; 1094 T; 0 other;

Query Match 41.1%; Score 453.2; DB 21; Length 4417;  
Best Local Similarity 90.1%; Pred. No. 6.1e-128;  
Matches 485; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 557 cacagctggccggattggagacagtaggaatcatgtttcccgtagtgacggcttcc 616  
Db 10 ctcaattggctggattgcagacagtaggaacacagtggttccgtggaacggcttcc 69  
Qy 617 aggcctgtcaaggcatccactcaatttgaagaagtcctgcacgaattctagcag 676  
Db 70 aggccttccaaggcttatccactcaatttgaagaagtcctgcacgaattctgccc 129  
Qy 677 ctgaaccatccactatgtcagagagagtcatttcaagttgctgagaggaacgct 736  
Db 130 ctgaaccatccactatgtcggggagagtccttcaagtggtcgtgaaggaagtgct 189  
Qy 737 tcgcacatgccgcacatggacaggggtcctctgcagcttcaagatcaatgactggcca 796  
Db 190 tcgcacatgccgcacatggacaggggtcctctgcagcttcaagatcaatgactggcca 249  
Qy 797 cactcaatgagaagcccttttctgtggaagacacacttatttactgtgccagccctatct 856  
Db 250 cgtcaatgagaagcccttttctgtggaagacacacttatttctgtgccagccacactct 309  
Qy 857 taaaagaagttggcatgaaagctcaactccaggtcagcatgaacatggcctctcttta 916  
Db 310 tgaagaagttggcatgaaagctcaactgcaggtcagcatgaacagcgcctgtccttca 369  
Qy 917 tctcaggttctgtcatcatcaccacacacactgttctgaagggttccatcctggccatcg 976  
Db 370 tctcaggttctgtcatcatcaccacacacactgttctgaagggtcctcctcctggcgattg 429  
Qy 977 cctctgtgatcctgttctgtcctctagccctgtctctctctgtggttctggccctct 1036  
Db 430 cctctgtgtgtctctctgtctgtggttctggcctgtggttctggccctctct 489  
Qy 1037 gctgcagctgtgattatcaaggaggttccctccaccctcctgcagagagagtgaggaata 1094  
Db 490 gctgcagctgtgattatcaaggaggttccctccaccctcctgttgagagagtgaggaata 547

RESULT 8  
AAL19043  
ID AAL19043 standard; cDNA; 356 BP.  
XX  
AC AAL19043;  
XX

DT 07-DEC-2001 (first entry)  
XX  
XX Human breast cancer expressed polynucleotide 11500.  
XX  
XX Human: breast cancer; cell marker; cytostatic; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200151628-A2.  
XX  
XX 19-JUL-2001.  
XX  
XX 10-JAN-2001; 2001WO-US00798.  
XX  
XX 14-JAN-2000; 2000US-0176077.  
XX  
XX 14-MAR-2000; 2000US-0189167.  
XX  
XX 24-MAR-2000; 2000US-0192099.  
XX  
XX 29-MAR-2000; 2000US-0193480.  
XX  
XX 15-MAY-2000; 2000US-0205230.  
XX  
XX 09-JUN-2000; 2000US-0211315.  
XX  
XX 25-JUL-2000; 2000US-0220534.  
XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
XX Lillie J, Xu Y, Wang Y, Steinmann K;  
XX  
XX WPI; 2001-451856/48.  
XX  
XX New peptide useful as a marker for the diagnosis of breast cancer -  
XX  
XX Claim 1; Page 2056; 3695pp; English.  
XX  
XX The invention relates to human breast cancer expressed polynucleotides  
CC (AAL07544-AAL26789) and methods of assessing whether a patient is  
CC afflicted with breast cancer by examining the correlation between the  
CC expression of certain markers and the cancerous state of breast cells.  
CC The polynucleotides and encoded polypeptides are potential markers for  
CC detecting, diagnosing, monitoring, characterizing treating and  
CC potentially preventing breast cancer. The polynucleotides and encoded  
CC polypeptides are also useful for isolating compounds with cytostatic  
CC activity.  
XX  
XX SQ Sequence 356 BP; 91 A; 78 C; 95 G; 92 T; 0 other;

Query Match 31.6%; Score 349; DB 22; Length 356;  
Best Local Similarity 100.0%; Pred. No. 1.3e-96;  
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 413 ggtacagacagccagcgcgtcatcattgttgcactgagtgaggaactccatgaagatctct 472  
Db 8 ggtacagacagccagcgcgtcatcattgttgcactgagtgaggaactccatgaagatctct 67  
Qy 473 tttctatttcagagagggaggttaagtctcagactcttggtgcaattgtttactgtg 532  
Db 68 tttctatttcagagagggaggttaagtctcagactcttggtgcaattgtttactgtg 127  
Qy 533 ttggtgtgaagatttcaatgagacacagctggcccggtatgcggacagtaggagatcatg 592  
Db 128 ttggtgtgaagatttcaatgagacacagctggcccggtatgcggacagtaggagatcatg 187  
Qy 593 tgtttcccgtagaagcagcgttccaggtctcgaaggatcatccactcaattttgaaga 652  
Db 188 tgtttcccgtagaagcagcgttccaggtctcgaaggatcatccactcaattttgaaga 247  
Qy 653 agtctcgtatcgaattctagcagctgaacctcaccatattgtgcaggagagatcttcc 712  
Db 248 agtctcgtatcgaattctagcagctgaacctcaccatattgtgcaggagagatcttcc 307  
Qy 713 aagtgtcgtgagaggaacggttcccgacatgcccgcaacggtggacag 761  
Db 308 aagtgtcgtgagaggaacggttcccgacatgcccgcaacggtggacag 356

```
RESULT 9
AAL08133
ID AAL08133 standard; cDNA; 355 BP.
XX AC
XX AAL08133;
XX DT 07-DEC-2001 (first entry)
XX DE Human breast cancer expressed polynucleotide 590.
XX KW Human; breast cancer; cell marker; cytostatic; ss.
XX OS Homo sapiens.
XX PN WO200151628-A2.
XX PD 19-JUL-2001.
XX PF 10-JAN-2001; 2001WO-US00798.
XX PR 14-JAN-2000; 2000US-0176077.
XX PR 14-MAR-2000; 2000US-0189167.
XX PR 24-MAR-2000; 2000US-0192099.
XX PR 29-MAR-2000; 2000US-0193480.
XX PR 15-MAY-2000; 2000US-0205230.
XX PR 09-JUN-2000; 2000US-0211315.
XX PR 25-JUL-2000; 2000US-0220534.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX DR WPI; 2001-451856/48.
XX PT New peptide useful as a marker for the diagnosis of breast cancer -
XX PS Claim 1; Page 189; 3695pp; English.
XX CC The invention relates to human breast cancer expressed polynucleotides
XX CC (AAL07544-AAL26789) and methods of assessing whether a patient is
XX CC afflicted with breast cancer by examining the correlation between the
XX CC expression of certain markers and the cancerous state of breast cells.
XX CC The polynucleotides and encoded polypeptides are potential markers for
XX CC detecting, diagnosing, monitoring, characterising treating and
XX CC potentially preventing breast cancer. The polynucleotides and encoded
XX CC polypeptides are also useful for isolating compounds with cytostatic
XX CC activity.
XX SQ Sequence 355 BP; 91 A; 78 C; 94 G; 92 T; 0 other;

Query Match 31.5%; Score 348; DB 22; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.5e-96;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 ggtacagagcagcgcgtcatctgttctgactgatggagaactccatgaagatctct 472
DB 8 ggtacagagcagcgcgtcatctgttctgactgatggagaactccatgaagatctct 67
QY 473 ttcttattcagagagggagcctaataagttctcgagatcttggcgcaattgtttactgtg 532
DB 68 ttcttattcagagagggagcctaataagttctcgagatcttggcgcaattgtttactgtg 127
QY 533 ttggtgtgaagatttcaatgagacacagctggcccgagattgcggacagtaaggatcatg 592
DB 128 ttggtgtgaagatttcaatgagacacagctggcccgagattgcggacagtaaggatcatg 187
QY 593 tgtttcccggtgaatgacggcttctcagagctctgcaaggcatcaccatcaattttgaaga 652
DB 188 tgtttcccggtgaatgacggcttctcagagctctgcaaggcatcaccatcaattttgaaga 247
QY 653 agtctgcatcgaattctagcagctgaaccatccaccatcatgtgcaggagagtcatttc 712

|||||
Db 248 agtctgcatcgaattcttagcagctgaaccatccaccatcatgtgcaggagagtcatttc 307
QY 713 aagttgtcgtgagagaaacggcttccgacatgcccgcaacgtggaca 760
Db 308 aagttgtcgtgagagaaacggcttccgacatgcccgcaacgtggaca 355

RESULT 10
AAL25878
ID AAL25878 standard; cDNA; 346 BP.
XX AC
XX AAL25878;
XX DT 07-DEC-2001 (first entry)
XX DE Human breast cancer expressed polynucleotide 18335.
XX KW Human; breast cancer; cell marker; cytostatic; ss.
XX OS Homo sapiens.
XX PN WO200151628-A2.
XX PD 19-JUL-2001.
XX PF 10-JAN-2001; 2001WO-US00798.
XX PR 14-JAN-2000; 2000US-0176077.
XX PR 14-MAR-2000; 2000US-0189167.
XX PR 24-MAR-2000; 2000US-0192099.
XX PR 29-MAR-2000; 2000US-0193480.
XX PR 15-MAY-2000; 2000US-0205230.
XX PR 09-JUN-2000; 2000US-0211315.
XX PR 25-JUL-2000; 2000US-0220534.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX DR WPI; 2001-451856/48.
XX PT New peptide useful as a marker for the diagnosis of breast cancer -
XX PS Claim 1; Page 3388; 3695pp; English.
XX CC The invention relates to human breast cancer expressed polynucleotides
XX CC (AAL07544-AAL26789) and methods of assessing whether a patient is
XX CC afflicted with breast cancer by examining the correlation between the
XX CC expression of certain markers and the cancerous state of breast cells.
XX CC The polynucleotides and encoded polypeptides are potential markers for
XX CC detecting, diagnosing, monitoring, characterising treating and
XX CC potentially preventing breast cancer. The polynucleotides and encoded
XX CC polypeptides are also useful for isolating compounds with cytostatic
XX CC activity.
XX SQ Sequence 346 BP; 90 A; 75 C; 90 G; 91 T; 0 other;

Query Match 31.3%; Score 346; DB 22; Length 346;
Best Local Similarity 100.0%; Pred. No. 1e-95;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 416 acaggacagcagcgcgtcatctgttctgactgatggagaactccatgaagatctcttt 475
DB 1 acaggacagcagcgcgtcatctgttctgactgatggagaactccatgaagatctcttt 60
QY 476 tcatttcagagagggagcctaataagttctcgagatcttggcgcaattgtttactgttg 535
DB 61 tcatttcagagagggagcctaataagttctcgagatcttggcgcaattgtttactgttg 120
QY 536 gttgtgaagatttcaatgagacacagctggcccgagattgcggacagtaaggatcatgtg 595
|||||
```

Db 121 gtgtgaagatttcaatgagacacagctggcccgattgcgagacagtaaggatctgtg 180  
Qy 596 ttcccgtaatagcagcgttcaagctctgcaaggcatcatcaatcttttgaagaagt 655  
Db 181 ttcccgtaatagcagcgttcaagctctgcaaggcatcatcaatcttttgaagaagt 240  
Qy 656 cctgcatacaattctagcagctgaaccatccaccatatgtgcagagagtcatttcaag 715  
Db 241 cctgcatacaattctagcagctgaaccatccaccatatgtgcagagagtcatttcaag 300  
Qy 716 ttgtcgtgagagaaacggctccgcacatgcccgcacaagtgagacg 761  
Db 301 ttgtcgtgagagaaacggctccgcacatgcccgcacaagtgagacg 346  
RESULT 11  
ID AAA75149 standard; cDNA; 3677 BP.  
XX  
AC AAA75149;  
XX  
DT 15-JAN-2001 (first entry)  
XX  
DE cDNA encoding a human TANGO 216 polypeptide.  
XX  
KW TANGO 266; TANGO 216; TANGO 262; TANGO 267;  
KW cellular proliferation; cellular differentiation; cellular adhesion;  
KW von Willebrand factor-associated disorder; cell trafficking; cancer;  
KW hematopoietic associated disease; atelectasis; pulmonary congestion;  
KW oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;  
KW intestinal disorder; spleen associated disease; renal disorder;  
KW cardiovascular disorder; ischemic heart disease; hydrocephalus;  
KW brain herniation; iatrogenic disease; inflammation; meningitis;  
KW Alzheimer's Disease; cerebral toxoplasmosis; Parkinson's disease;  
KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 307..1773  
FT /\*tag= a  
FT /product= "TANGO 216"  
FT sig\_peptide 307..405  
FT /\*tag= b  
FT mat\_peptide 406..1773  
FT /\*tag= c  
XX  
PN WO200052022-A1.  
XX  
PD 08-SEP-2000.  
XX  
PF 01-MAR-2000; 2000WO-US05226.  
XX  
PR 01-MAR-1999; 99US-0122458.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Barnes TM, Holtzman DA, Sharp JD, Fraser CC;  
XX  
XX WPI: 2000-579269/54.  
DR P-PSDB; AAB18447.  
XX  
XX Novel human and murine secreted proteins designated TANGO 216, 261,  
PT 262, 266 and 267 useful as modulating agents of cellular processes,  
PT e.g. for treating cancer -  
XX  
PS Claim 2; Fig 1A-C; 175pp; English.  
XX  
CC The present sequence encodes a human TANGO 216 polypeptide. The  
CC specification also describes TANGO 266, TANGO 261, TANGO 262, and  
CC TANGO 267. The TANGO polypeptides can be used to modulate cellular  
CC proliferation, modulate cellular differentiation and/or modulate  
CC cellular adhesion. The proteins can be used to treat any von Willebrand

CC factor-associated disorder, regulate extracellular matrix structuring,  
CC cellular adhesion, and cell trafficking and/or migration, modulate  
CC cellular interactions, modulate cell adhesion in proliferative  
CC disorders, such as cancer, modulate the proliferation, differentiation,  
CC and/or function of cells that appear in the bone marrow, and leukocytes,  
CC treat bone marrow, blood and hematopoietic associated diseases and  
CC disorders, atelectasis, pulmonary congestion or oedema, emphysema,  
CC chronic bronchitis, bronchial asthma and bronchiectasis, intestinal  
CC disorders, spleen associated diseases, modulate renal disorders, treat  
CC cardiovascular disorders such as ischemic heart disease, modulate the  
CC proliferation, differentiation, and/or function of bone and cartilage  
CC cells and to treat bone and/or cartilage associated diseases or  
CC disorder. They may also be used to treat disorders associated with the  
CC ovaries, cerebral oedema, hydrocephalus, brain herniations, iatrogenic  
CC diseases, inflammations, bacterial and viral meningitis, Alzheimer's  
CC disease, cerebral toxoplasmosis, Parkinson's disease, multiple sclerosis,  
CC brain cancers, hydrocephalus and encephalitis, and treat hepatic  
CC disorders.  
XX  
SQ Sequence 3677 BP; 1055 A; 776 C; 881 G; 965 T; 0 other;  
Query Match 31.2%; Score 344.6; DB 21; Length 3677;  
Best Local Similarity 58.2%; Pred. No. 1.1e-94;  
Matches 627; Conservative 0; Mismatches 444; Indels 6; Gaps 1;  
Qy 1 atggccacggcgagcgagagcgccctcgcatcggtctccagtgctctttggccact 60  
Db 307 atggcgcgagcggttcccgccgcgcagcccgagctggtcttcccggctgtgg 366  
Qy 61 ctgggtgctcatctgcgcggcgcaaggcgagggagtggtgggttcacgcctgtac 120  
Db 367 ctgtgtgtcagcggttcccgggggtgctgcgcgccagggagcagccctctgcaga 426  
Qy 121 gcggatttgactctactcttcttgacaaatcaggaagtgtgctgcaccactggaat 180  
Db 427 agagcctttgactctactcttctgcagaaagtgtggagtggtggcaataactggatt 486  
Qy 181 gaaatctattactttgtggaacagttggctcacaaattcatcagccacagttggaatg 240  
Db 487 gaaattataattctgcacagcaacttgcggagagatttgcagccctgaaatgagatta 546  
Qy 241 tctttattgttttctccaccgcaggaacaaacttaataaactgacagagacagagaa 300  
Db 547 tcttcaattgttttcttctcaagcaactattatttggccatttaactggagacagagc 606  
Qy 301 caaatcctcgaaggcctagaagaactccagaagaacttctgcagggagagacacttacatg 360  
Db 607 aaaatcagtaaaaggcttgagagatttaaaacgtgttagtccagtaggagacatatc 666  
Qy 361 catgaaggatttgaaaggcgccagtgagcagatttatatgaaaaacagacaagggtacagg 420  
Db 667 catgaaggactaaagctagcgaatgaacaaat-----tcagaaagcaggaggttgaaa 720  
Qy 421 acagccagcgtcatctattgttgcagctgatgagaaactccatgaaagatctctttctat 480  
Db 721 acctcagtcataatgtctgcagatggcaagttggaagttggcgtctgtgtgcatacat 780  
Qy 481 tcagagagggaggcttaattaggtctgcagatcttggcgcaattgtttactgtgtgtgtg 540  
Db 781 gcagagaaagaggcaagatatccagtcacttggggcagtgatttattgtgtgtgtgtgc 840  
Qy 541 aaagatttcaatgagacagagctggcccgattgcggagtgcgagcagtaagatcatgtttccc 600  
Db 841 cttgattttgaacagcacagcttgaaagaattgctgattcccaagagagaagtttccct 900  
Qy 601 gtgaatgacgctttcaggtctctgcaagcctacatccactcaattttgaaagtcctgc 660  
Db 901 gtcaaggctggatttcaggctctttaaaggaaataataattcttactactagctcagtcagt 960  
Qy 661 atcgaaattctagcagctgaaccatccaccatatgtgcagggagagtcatttcaagtgtgc 720  
Db 961 actgaaatcctagaattgcagccctcaagtgtctgtgtggggaggagaatttcagattgc 1020







QY 481 tcagagaggaggagtaaggctcagagatcttggtgcaattgtttactgtgtgtg 540  
DB 781 gcagagaaagaggcaagatattcaggtcacttgggctagtggttctgtgtggtc 840  
QY 541 aaagatttcaatgagacacagctggccgagtgagcagtaagatgatgtgttccc 600  
DB 841 cttgattttgaacaagcacagcttgaaagaattgctgattccaaaggagcaagtttccct 900  
QY 601 gtgaatgacggcttttcaggtctgaaggcatcatccactcaatttttgaagaagtcctgc 660  
DB 901 gtcaaggttgatttcaggtctttaaaggaaataattctatactactcagtcagtcgt 960  
QY 661 atcgaattctagcagctgaaccatccatattgtgcaggagagtcatttcaagttgtc 720  
DB 961 actgaatcctagaattgcagccctcaagtgtctgtgtggggaggaatttcagattgtc 1020  
QY 721 gtgagaggaacggcttcagacatgcccgcaacgtggagcagggctctctgcagcttcaag 780  
DB 1021 ttaagtgaagaggattcatgctggcgagtcggaatggcagtggttctctgcacttacct 1080  
QY 781 atcaatgactcgtcacactcaatgagaagcccttttctgtgtggaagacacttattactg 840  
DB 1081 gtaaatgaacatatacaacgagtgtaaacacagtaagtgtacagcttaattctatgctt 1140  
QY 841 gtccagcgctctatctaaagaagtggcatgaaagctgcactccaggtcagcatgaac 900  
DB 1141 tctctgcactatctgtaataaagctggagaaactctgtgatttcagtgagctttaat 1200  
QY 901 gatggcctcttttctccagttctgtctatcatcacaccacacacactgtttcagcgt 960  
DB 1201 ggaagaaatctgtcttcttcaggtattcaattgtcacagccacagaaatgttcaacggg 1260  
QY 961 tcatcctggcactgcctgctgactgttctctgtctctactagccctggtctctctg 1020  
DB 1261 atcgagccatcatgttatttgggtgactgctactctctgggagctgtgtgagtg 1320  
QY 1021 tggttctgccccctctgctgactgtgattatcaaggaggtccctccacccccctgc 1077  
DB 1321 tggtttggcccccttgcgcaagtggtatttaaggatcctccaccaccccccc 1377

RESULT 14  
AAA75159  
ID AAA75159 standard; cDNA; 3677 BP.  
XX  
AC AAA75159;  
XX  
DT 15-JAN-2001 (first entry)  
XX  
DE cDNA clone encoding a human TANGO 216 polypeptide.  
XX  
KW TANGO 266; TANGO 216;; TANGO 261; TANGO 262; TANGO 267;  
KW cellular proliferation; cellular differentiation; cellular adhesion;  
KW von Willebrand factor-associated disorder; cell trafficking; cancer;  
KW hematopoietic associated disease; atelectasis; pulmonary congestion;  
KW oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;  
KW intestinal disorder; spleen associated disease; renal disorder;  
KW cardiovascular disorder; ischemic heart disease; hydrocephalus;  
KW brain herniation; iatrogenic disease; inflammation; meningitis;  
KW Alzheimer's Disease; cerebral toxoplasmosis; Parkinson's disease;  
KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
FH Location/Qualifiers  
CDS 307..1773  
FT /\*tag= a  
FT /product= "TANGO 216"  
XX  
PN WO200052022-A1.  
XX  
PD 08-SEP-2000.

XX  
PF 01-MAR-2000; 2000WO-US05226.  
XX  
PR 01-MAR-1999; 99US-0122458.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Barnes TM, Holtzman DA, Sharp JD, Fraser CC;  
XX  
DR WPI: 2000-579269/54.  
DR P-PSDB; AAB18457.  
XX  
PT Novel human and murine secreted proteins designated TANGO 216, 261,  
PT 262, 266 and 267 useful as modulating agents of cellular processes,  
PT e.g. for treating cancer -  
XX  
PS Disclosure; Page -: 175pp; English.  
XX  
CC AAA75157-59 encode human TANGO 216 proteins. The specification also  
CC describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO  
CC polypeptides can be used to modulate cellular proliferation, modulate  
CC cellular differentiation and/or modulate cellular adhesion. The  
CC proteins can be used to treat any von Willebrand factor-associated  
CC disorder, regulate extracellular matrix structuring, cellular adhesion,  
CC and cell trafficking and/or migration, modulate cellular interactions,  
CC modulate cell adhesion in proliferative disorders, such as cancer,  
CC modulate the proliferation, differentiation, and/or function of cells  
CC that appear in the bone marrow, and leukocytes, treat bone marrow, blood  
CC and hematopoietic associated diseases and disorders, atelectasis,  
CC pulmonary congestion or oedema, emphysema, chronic bronchitis, bronchial  
CC asthma and bronchiectasis, intestinal disorders, spleen associated  
CC diseases, modulate renal disorders, treat cardiovascular disorders such  
CC as ischemic heart disease, modulate the proliferation, differentiation,  
CC and/or function of bone and cartilage cells and to treat bone and/or  
CC cartilage associated diseases or disorder. They may also be used to  
CC treat disorders associated with the ovaries, and cerebral oedema,  
CC hydrocephalus, brain herniations, iatrogenic disease, inflammations,  
CC bacterial and viral meningitis, Alzheimer's disease, cerebral  
CC toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers,  
CC hydrocephalus and encephalitis, and treat hepatic disorders.  
CC note: the present sequence does not appear in the specification; it was  
CC created using information provided.  
XX  
SQ Sequence 3677 BP; 1054 A; 777 C; 881 G; 965 T; 0 other;

Query Match 31.1%; Score 343; DB 21; Length 3677;  
Best Local Similarity 58.1%; Pred. No. 3.5e-94;  
Matches 626; Conservative 0; Mismatches 445; Indels 6; Gaps 1;  
QY 1 atggccacggcgagcgagagccctcgccgcatcgcttccagtggtctcttggccact 60  
DB 307 atggtggcgagcggtcccgcccgccgagcgctgtgttcccgggctgtgg 366  
QY 61 ctgtgtctcatctgcgcggcgcaaggggcgagcgagagatggggtccagcctctac 120  
DB 367 ctgtgtggtctcagcgtcccgggggtgctgcgcgcccgagcagcctctcgaga 426  
QY 121 ggcggattgacctgtactctatttggacaatcaggaaagtgtgtgcaccactggaat 180  
DB 427 agagccttgatctctactctcctcgacaagctgggagtggtggcaataactggatt 486  
QY 181 gaaatctattcttggaaacagttggctcacaattctcatcgccacagttgagaatg 240  
DB 487 gacattataattcttcacagcaacttgcgagagatttgggcccctgaaagagatta 546  
QY 241 tctttattgttttctccaccggaggaacaccttaataatgaaactgacagagacagaa 300  
DB 547 tcttctatgtgtttctctctcaagcaactattattttgcccattaaactgagacagagc 606  
QY 301 caaatccgtcaaggcctagaagaactccagaaagtcttgcaggaggagacattacatg 360  
DB 607 aaatcagtaaaaggcttggaggatttaaaacgtgtttagtccagtaggagagacata 666

QY 361 catgaagatttgaagggccagtgagcagatttattatgaacacagacaggggtacagg 420  
DB 667 catgaagactaaagctgcggaatgaacaaat-----tcagaagcgaggagcttgaaa 720  
QY 421 acagcagcgtcatctgcttctgactgatgagaaactccatgaagatctcttttctat 480  
DB 721 aactcagtatcataatgctgctgacagatgcaagttggagcgtctggtgccatcatat 780  
QY 481 tcagagagggagcctaagtgctcagagatcttggtgcaaatggttactgtgttggtg 540  
DB 781 gcagagaaagggcaagatatccaggctcaacttgggctggttattgttgggtg 840  
QY 541 aaagatttcaatgagacacagctgccgagttgcgacagtaagatcatgtgtttcc 600  
DB 841 cttgattttgaacagcagctgaaagaaatctgctgattccaaaggagcaagtttccct 900  
QY 601 gtgaatgacgctttcaggtctgcaagcctcatccatccactcaattttgaagaagtcctgc 660  
DB 901 gtcaaggtggatttcaggctcttaagaaataaataattctatactagctcagtcagt 960  
QY 661 atcgaattctcagcagtgaaacccatccatatagtgagggagagtcatttcaagttg 720  
DB 961 actgaaatcctagaattgagccctcaagtgctgtgtgggggaggaatttcagattg 1020  
QY 721 gtgagaggaacgctctccgcacatgcccgcaacgtggacaggtcctctgcagctcaag 780  
DB 1021 ttaagtggagaggattctatctgtggcagtcggaatgagcagtgcttctgcactacact 1080  
QY 781 atcaatgactcgtgcacactcaatgagaagccctttctgtggaagacacttattactg 840  
DB 1081 gtaaatgaacatatcaacagcagtgtaaaacacagtaagtgtacagcttaattctatg 1140  
QY 841 ttctcagcgcctatttataaaagaagtggcagtgaaagtcgactccaggtcagcatgaac 900  
DB 1141 ttctcgtgcactatcctgaaatgagaaactctgtgttccagtgagctttaa 1200  
QY 901 gatggcctctttatctccagttctgtcatcatcacacacacacactgttctgacgt 960  
DB 1201 gaggaataatcgtcatttcagatcataatgtcacagcacagaatgttctcaagg 1260  
QY 961 tccactcgtgcacgcctgctgctgctgctgctgctgctgctgctgctgctgctg 1020  
DB 1261 atcgcagcactatgttatttgggttactgctgctgctgctgctgctgctgctg 1320  
QY 1021 tggttcggccctctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1077  
DB 1321 tggtttggcccttctgctgcaagtggttataaggatcctccaccaccccccc 1377

RESULT 15  
AAA75162  
ID AAA75162 standard; cDNA; 3501 BP.  
XX  
AC AAA75162;  
XX  
DT 15-JAN-2001 (first entry)  
XX  
DE cDNA clone encoding a murine TANGO 216 polypeptide.  
XX  
KW TANGO 266; TANGO 216;; TANGO 261; TANGO 262; TANGO 267;  
KW cellular proliferation; cellular differentiation; cellular adhesion;  
KW von Willebrand factor-associated disorder; cell trafficking; cancer;  
KW hematopoietic associated disease; atelectasis; pulmonary congestion;  
KW oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;  
KW intestinal disorder; spleen associated disease; renal disorder;  
KW cardiovascular disorder; ischemic heart disease; hydrocephalus;  
KW brain herniation; iatrogenic disease; inflammation; meningitis;  
KW Alzheimer's Disease; cerebral toxoplasmosis; Parkinson's disease;  
KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder; ss.  
XX  
OS Mus sp.  
XX

FH Key Location/Qualifiers  
FT CDS 149..1612  
FT /\*tag= a  
FT /product= "TANGO 216"  
XX  
\* WO200052022-A1.  
PN 08-SEP-2000.  
XX  
PF 01-MAR-2000; 2000WO-US052226.  
XX  
XX 01-MAR-1999; 99US-0122458.  
PR (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Barnes TM, Holtzman DA, Sharp JD, Fraser CC;  
XX  
XX WPI; 2000-579269/54.  
DR P-PSDB; AAB18460.  
XX  
PT Novel human and murine secreted proteins designated TANGO 216, 261,  
PT 262, 266 and 267 useful as modulating agents of cellular processes,  
PT e.g. for treating cancer -  
XX  
PS Disclosure; Page -: 175pp; English.  
XX  
XX AAA75160-62 encode murine TANGO 216 proteins. The specification also  
CC describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO  
CC polypeptides can be used to modulate cellular proliferation, modulate  
CC cellular differentiation and/or modulate cellular adhesion. The  
CC proteins can be used to treat any von Willebrand factor-associated  
CC disorder, regulate extracellular matrix structuring, cellular adhesion,  
CC and cell trafficking and/or migration, modulate cellular interactions,  
CC modulate cell adhesion in proliferative disorders, such as cancer,  
CC modulate the proliferation, differentiation, and/or function of cells  
CC that appear in the bone marrow, and leukocytes, treat bone marrow, blood  
CC and hematopoietic associated diseases and disorders, atelectasis,  
CC pulmonary congestion or oedema, emphysema, chronic bronchitis, bronchial  
CC asthma and bronchiectasis, intestinal disorders, spleen associated  
CC diseases, modulate renal disorders, treat cardiovascular disorders such  
CC as ischemic heart disease, modulate the proliferation, differentiation,  
CC and/or function of bone and cartilage cells and to treat bone and/or  
CC cartilage associated diseases or disorder. They may also be used to  
CC treat disorders associated with the ovaries, and cerebral oedema,  
CC hydrocephalus, brain herniations, iatrogenic disease, inflammations,  
CC bacterial and viral meningitis, Alzheimer's Disease, cerebral  
CC toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers,  
CC hydrocephalus and encephalitis, and treat hepatic disorders.  
CC note: the present sequence does not appear in the specification; it was  
CC created using information provided.  
XX  
XX Sequence 3501 BP; 974 A; 795 C; 833 G; 899 T; 0 other;

Query Match 31.0%; Score 341.8; DB 21; Length 3501;  
Best Local Similarity 60.0%; Pred. No. 7.9e-94;  
Matches 590; Conservative 0; Mismatches 387; Indels 6; Gaps 1;  
QY 109 ccagcctgctacgcgagatttgacctgacttctgacaaatccaggaagtgtgctg 168  
DB 257 cctcttgcaaaaagccttcgattgtactgtactggacaaagtctggcagtgtagca 316  
QY 169 caccactggaatgaatcttacttcttggacagttggtcacaattcatcagccca 228  
DB 317 aataactggtatgaaattataattttgtccacagctgacagacagatttggagcct 376  
QY 229 cagttgagaatgctcttattgttttccaccgaggaacaccttaataaactgaca 288  
DB 377 gaaatgagattgtctctcattgttttctcccaagcaaccattattttgcoattca 436  
QY 289 gaagacagagacaaatccgtcaaggcctcagaaactccagaaagtcttgcagaggga 348  
DB 437 gagacaggtacaaaattggcaaggactggagatttaaaagccggttaagccagtgg 496

Qy 349 gacacttacatgcataagagatttgaaaggccagtgagcagatttattatgaaacaga 408  
 Db 497 gaaacatacatccatgaagactaaagcttgcaaacgacaaat-----tcaaaatgca 550  
 Qy 409 caagggtacagagacagcagcgtcatcatctgttctgactgatggagaactccatgaagat 468  
 Db 551 ggaggcttaaaagctccagtatcataattgctttgacggacggtaaagctggacggcctg 610  
 Qy 469 ctcttttcttctacagagagagaggaactaaataggctcagatcttctggaattgtttac 528  
 Db 611 gtacatcttatgagagaaagaggaagaaagtcacaggtcacttggcgtagtgttac 670  
 Qy 529 tgtgtgtgtgaaagatttcaatgagacacagctggcccgatttgcggacagtaaggat 588  
 Db 671 tgcgttggggctcttgattttgaacaaagctcagctggaaagaattgctgattccaaggac 730  
 Qy 589 catgtgttcccgtaatagcaggttttcagcgtctcgaaggcatcatccatcaattttg 648  
 Db 731 cagggtttccctgtcaaaaggtggatttcaagctctcaaaagcatcatcaactctatatta 790  
 Qy 649 aagaagtcctgcatacgaataattctagcagctgaaccatccaccatattgtcaggagagca 708  
 Db 791 gctcaatcatgtactgaaatcctggaattgagtccttcaagtgtctgtgtagggagagaaa 850  
 Qy 709 ttcaagttgtcgtgagagaaagcgttccgcacatgcccgcacagtggaagggtcctc 768  
 Db 851 ttccaagttgtcgtgactggaagagcagtcagtcagtcagtcagtcagtcagtcagtc 910  
 Qy 769 tgcaggttcaagatcaatgactcgtgcacactcaatgagaagcccttttctgtggaagac 828  
 Db 911 tgtacattcactgcaaacagcacatatacaaaagagtgagagccagtgagcattcagcca 970  
 Qy 829 acttatttactgtgtccagcgccttatcttaaaagaagttggcatgaaagctgcactccag 888  
 Db 971 agttccatcccttgcctgcacctgtcctgacaaagatggagaacctcttgaagtcca 1030  
 Qy 889 gtcagcatgacgattggcctctcttttctccagttctgtcatcatcaccaccacacac 948  
 Db 1031 atcagctataatgatgggaagctgctgtctcaagatccttaacaatcacagccacagaa 1090  
 Qy 949 tgttctgacggttccatccttgccatcgccctgtgatacctgttctcctctagccctg 1008  
 Db 1091 tgtaccaatggattgcagccatcgtagctatttttgggtgtgtgtgtgtgtgtgtgtgt 1150  
 Qy 1009 gctctcctctggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1068  
 Db 1151 gccttgaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1210  
 Qy 1069 cccctgcgagagagagtgagga 1091  
 Db 1211 ccaccttctgcaccaatggagga 1233

**THIS PAGE BLANK (uspto)**



```

; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/308,829
; PRIOR FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 359
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 231
; LENGTH: 5540
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-08253-231

Query Match      84.4%; Score 1193.8; DB 1; Length 5540;
Best Local Similarity 99.8%; Pred. No. 9.1e-160;
Matches 1195; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aggaccccgagggaagggcccgcgatggcgctccctcgagggtcggtgagagttcgcg 60
Db 41 aggaccccgagggaagggcccgcgatggcgctccctcgagggtcggtgagagttcgcg 100
QY 61 agcgtgggaagagcgagccctgctctcccggtggtcggtgagagttcgcgagcg 120
Db 101 agcgtgggaagagcgagccctgctctcccggtggtcggtgagagttcgcgagcg 160
QY 121 gagagccctcgcatcggtctccagtggctctcttggccactctggtgctcatctgcg 180
Db 161 gagagccctcgcatcggtctccagtggctctcttggccactctggtgctcatctgcg 220
QY 181 cgggcaaggagagcagagaggtgggtccagctgctacggtggtgatttgacctgta 240
Db 221 cgggcaaggagagcagagaggtgggtccagctgctacggtggtgatttgacctgta 280
QY 241 ctctatttggacaatcaggaagtgtgctgcacactgggaatgaaatctattacttgt 300
Db 281 ctctatttggacaatcaggaagtgtgctgcacactgggaatgaaatctattacttgt 340
QY 301 ggacagttggctcaaaaattcaagccacagttgagaaatgctcttattgtttctc 360
Db 341 ggacagttggctcaaaaattcaagccacagttgagaaatgctcttattgtttctc 400
QY 361 caccgaggaacacaccttaataagactgacagaagacagagaacaaatccgtcaaggcct 420
Db 401 caccgaggaacacaccttaataagactgacagaagacagagaacaaatccgtcaaggcct 460
QY 421 agaagaactccagaagaattctgcagagagagacacttacatgcatgaaggattgaaag 480
Db 461 agaagaactccagaagaattctgcagagagagacacttacatgcatgaaggattgaaag 520
QY 481 ggcagtgacagatttattatgaaacagacaaggtacagagcagcagcgctcatcat 540
Db 521 ggcagtgacagatttattatgaaacagacaaggtacagagcagcagcgctcatcat 580
QY 541 tgctttgactgatggagaactccatgaagatctcttttctatttcagagagaggtctaa 600
Db 581 tgctttgactgatggagaactccatgaagatctcttttctatttcagagagaggtctaa 640
QY 601 taggtctgagatcttggtgcaattgtttactgtgtgtggtgaaagatttcaatgagac 660
Db 641 taggtctgagatcttggtgcaattgtttactgtgtgtggtgaaagatttcaatgagac 700
QY 661 acagctggcccgattgcgagacagatcattgttttcttctatttcagagagaggtctaa 720
Db 701 acagctggcccgattgcgagacagatcattgttttcttctatttcagagagaggtctaa 760
QY 721 ggcctgcaaggcatccactcaattttgaaagaagtcctgcatcgaataatcttagcagc 780
Db 761 ggcctgcaaggcatccactcaattttgaaagaagtcctgcatcgaataatcttagcagc 820
QY 781 tgaaccatccaccatattgtagagagatcttcaagttgtcgtgagagaaacggctt 840
Db 821 tgaaccatccaccatattgtagagagatcttcaagttgtcgtgagagaaacggctt 880

RESULT 2
PCT-US02-08253-231
; Sequence 231, Application PC/TUS0208253
; GENERAL INFORMATION:
; APPLICANT: Carson-Walter, Eleanor
; APPLICANT: St. Croix, Brad
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00179
; CURRENT APPLICATION NUMBER: PCT/US02/08253
```

```
Qy 841 ccgacatcccgcaacgtggacaggttcctctgcagcttcaagatcaatgactcgggtcac 900
Db 881 ccgacatcccgcaacgtggacaggttcctctgcagcttcaagatcaatgactcgggtcac 940
Qy 901 actcaatgagaagccctttctgtggaagacactattattactgtgtccagcgccctatttt 960
Db 941 actcaatgagaagccctttctgtggaagacactattattactgtgtccagcgccctatttt 1000
Qy 961 aaagaagtggcatgaaagtgcactccaggttcagcgtgacgatgaacgatgcccctcttttat 1020
Db 1001 aaagaagtggcatgaaagtgcactccaggttcagcgtgacgatgaacgatgcccctcttttat 1060
Qy 1021 ctccagttctgtcatcatcacacacacactgttctgacggttccatccctgggccaatgc 1080
Db 1061 ctccagttctgtcatcatcacacacacactgttctgacggttccatccctgggccaatgc 1120
Qy 1081 cctgctgatcctgttctcctgctcctagcctggctcctctcctgtgtgttctggccctctg 1140
Db 1121 cctgctgatcctgttctcctgctcctagcctggctcctctcctgtgtgttctggccctctg 1180
Qy 1141 ctgcaactgtattatcaaggaggttcctccacccctcctccagggagagatgaggaaaaa 1197
Db 1181 ctgcaactgtattatcaaggaggttcctccacccctcctccagggagagatgaggaaaa 1237

RESULT 3
US-10-038-307-1
; Sequence 1, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038.307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2272
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (213)...(1211)
US-10-038-307-1
```

```
Query Match 74.4%; Score 1052.4; DB 7; Length 2272;
Best Local Similarity 99.9%; Pred. No. 7.7e-140;
Matches 1053; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 aggaccgcgaggaagggcccgagatggcgctccctgaggggtcgtgaggttcggttcgcgg 60
Db 110 aggaccgcgaggaagggcccgagatggcgctccctgaggggtcgtgaggttcggttcgcgg 169
Qy 61 agcgtgggaagagcgagccctgtctctcccggtcgtcgggccatggccacggcgagagcg 120
Db 170 agcgtgggaagagcgagccctgtctctcccggtcgtcgggccatggccacggcgagagcg 229
Qy 121 gagagcctcgcatcggttcacagtgcctctcttggccactctggtgctcatctgcgc 180
Db 230 gagagcctcgcatcggttcacagtgcctctcttggccactctggtgctcatctgcgc 289
Qy 181 cgggcaagggggacgcagggaggaagggttcacgctgctacgcggattgacctga 240
Db 290 cgggcaagggggacgcagggaggaagggttcacgctgctacgcggattgacctga 349
Qy 241 ctctattttggacaatcagggaagtgtgtgcaccactggaatgaatctattttgt 300
Db 350 ctctattttggacaatcagggaagtgtgtgcaccactggaatgaatctattttgt 409
```

```
Qy 301 ggaacagttggttcacaaattcatcagccacagttgagaatgtcctttattgttttttc 360
Db 410 ggaacagttggttcacaaattcatcagccacagttgagaatgtcctttattgttttttc 469
Qy 361 caccgaggaacaaccttaataatgaaactgacagaagacagagaacaatccgccaaggcct 420
Db 470 caccgaggaacaaccttaataatgaaactgacagaagacagagaacaatccgccaaggcct 529
Qy 421 agaagaactccagaaagtctgcaggaggagacacttacatgcatgaaggatttgaagaag 480
Db 530 agaagaactccagaaagtctgcaggaggagacacttacatgcatgaaggatttgaagaag 589
Qy 481 ggcagtgagcagatttattatgaaaaacagacaagggttacaggacagccagcgtcatcat 540
Db 590 ggcagtgagcagatttattatgaaaaacagacaagggttacaggacagccagcgtcatcat 649
Qy 541 tgccttgactgatggagaactccatgaagatctctttttctatttcagagagggagcttaa 600
Db 650 tgccttgactgatggagaactccatgaagatctcttttctatttcagagagggagcttaa 709
Qy 601 taggtctcgagatcttggtagaattgttttactgtgttgggtgaaagatttcaatgagac 660
Db 710 taggtctcgagatcttggtagaattgttttactgtgttgggtgaaagatttcaatgagac 769
Qy 661 acagctgcccggattcgagcagtaagatcatgtttccctgtgaatgacggttttca 720
Db 770 acagctgcccggattcgagcagtaagatcatgtttccctgtgaatgacggttttca 829
Qy 721 ggctgcaaggcatccatccactcaattttgaagaagtcctgcatcgaaattctagcagc 780
Db 830 ggctgcaaggcatccatccactcaattttgaagaagtcctgcatcgaaattctagcagc 889
Qy 781 tgaacatccaccatgatgtgcaggagagtcatttcaagtctgcgaggaagaaacggctt 840
Db 890 tgaacatccaccatgatgtgcaggagagtcatttcaagtctgcgaggaagaaacggctt 949
Qy 841 ccgacatcccgcaacgtggacaggttcctctgcagttcaagatcaatgactcgggtcac 900
Db 950 ccgacatcccgcaacgtggacaggttcctctgcagttcaagatcaatgactcgggtcac 1009
Qy 901 actcaatgagaagccctttctgtggaagacactattattactgtgtccagcgccctatt 960
Db 1010 actcaatgagaagccctttctgtggaagatacttattactgtgtccagcgccctatt 1069
Qy 961 aaagaagtggcatgaaagtgcactccaggttcagcgtgacgatgaacgatgcccctctttat 1020
Db 1070 aaagaagtggcatgaaagtgcactccaggttcagcgtgacgatgaacgatgcccctctttat 1129
Qy 1021 ctccagttctgtcatcatcaccaccacacactgt 1054
Db 1130 ctccagttctgtcatcatcaccaccacacactgt 1163

RESULT 4
US-10-038-307-17
; Sequence 17, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038.307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1674
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-038-307-17
```





Db 661 atcgaaattctcagctgaaccatccaccatattgtgcaggagagtcattttcaagtgtc 720  
Qy 824 gtgagagaaacgcttccgacatgccgcaacgtgacaggtcctctgcagctcaag 883  
Db 721 gtgagagaaacgcttccgacatgccgcaacgtgacaggtcctctgcagctcaag 780  
Qy 884 atcaatgactcgttcacactcaatgagaagcccttttctgtggaagacacttattactg 943  
Db 781 atcaatgactcgttcacactcaatgagaagcccttttctgtggaagatacttattactg 840  
Qy 944 tgtccagcgcctctctataaaagaagttggtgcatgaaagctgcactccaggtcagcatgaac 1003  
Db 841 tgtccagcgcctctctataaaagaagttggtgcatgaaagctgcactccaggtcagcatgaac 900  
Qy 1004 gatgacctcttttctcagttctcagttctgtcatcatcaccacacacactgttctgacggt 1063  
Db 901 gatgacctcttttctcagttctcagttctgtcatcatcaccacacacactgttctcccaaa 960  
Qy 1064 tc 1065  
Db 961 tc 962  
RESULT 6  
US-10-038-307-15  
; Sequence 15, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15  
; LENGTH: 1650  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-038-307-15

Query Match 67.4%; Score 952.4; DB 7; Length 1650;  
Best Local Similarity 99.4%; Pred. No. 9.3e-126;  
Matches 956; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Qy 104 atggccacgcgagcgagagcctcggcatcggtccagtggtctctttggccact 163  
Db 1 atggccacgcgagcgagagcctcggcatcggtccagtggtctctttggccact 60  
Qy 164 ctggtgctcatctgcgcgcgcaaggggagcgagggagagtggttccagcctgctac 223  
Db 61 ctggtgctcatctgcgcgcgcaaggggagcgagggagtggttccagcctgctac 120  
Qy 224 ggcggttgacctgactctattttgacaaatcaggaagtgtgctcacactggaaat 283  
Db 121 ggcggttgacctgactctattttgacaaatcaggaagtgtgctcacactggaaat 180  
Qy 284 gaaatctattactttgtggaacagttggtctcacaaatcagccacagttgagaatg 343  
Db 181 gaaatctattactttgtggaacagttggtctcacaaatcagccacagttgagaatg 240  
Qy 344 tcccttattgtttctccaccggaggaacaccttaataagaactgacagaagagaaa 403  
Db 241 tcccttattgtttctccaccggaggaacaccttaataagaactgacagaagagaaa 300  
Qy 404 caaatccgtcaagcctagaagaacctcagaaaagttctgcaggagagacacttcacatg 463  
Db 301 caaatccgtcaagcctagaagaacctcagaaaagttctgcaggagagacacttcacatg 360  
Qy 464 catgaaggatttgaaggggccagtgagcagattattatgaaaaacagacaagggttacagg 523

Db 361 catgaaggatttgaaggggccagtgagcagatttattatgaaaaacagacaagggttacagg 420  
Qy 524 acagccagcgctcatcattgctttgactgatggagaacctccatgaagatctcttttctcat 583  
Db 421 acagccagcgctcatcattgctttgactgatggagaacctccatgaagatctcttttctcat 480  
Qy 584 tcagagagggagggctaaataggtctcgcagatcttggtgcaattgtttactgtgtggtgtg 643  
Db 481 tcagagagggagggctaaataggtctcgcagatcttggtgcaattgtttactgtgtggtgtg 540  
Qy 644 aaagatttcaatgagacacagctggcccgattgcggagatgcggagatcaatgtgtttccc 703  
Db 541 aaagatttcaatgagacacagctggcccgattgcggagatcaatgtgtttccc 600  
Qy 704 gtgaatgacggtcttccaggtctgcaaggcatccactcaattttgagaagcctctgc 763  
Db 601 gtgaatgacggtcttccaggtctgcaaggcatccactcaattttgagaagcctctgc 660  
Qy 764 atcgaaattctagcagctgaaccatccaccatattgtgcaggagagctatttcaagtgtc 823  
Db 661 atcgaaattctagcagctgaaccatccaccatattgtgcaggagagctatttcaagtgtc 720  
Qy 824 gtgagagaaacggtcttccgacatgccgcaacgtggagaggggtcctctgcagcttcaag 883  
Db 721 gtgagagaaacggtcttccgacatgccgcaacgtggagaggggtcctctgcagcttcaag 780  
Qy 884 atcaatgactcgttcacactcaatgagaagcccttttctgtggaagacacttattactg 943  
Db 781 atcaatgactcgttcacactcaatgagaagcccttttctgtggaagactatttactg 840  
Qy 944 tgtccagcgcctattttaaagaagttggcatgaaagctgcactccaggtcagcatgaac 1003  
Db 841 tgtccagcgcctattttaaagaagttggcatgaaagctgcactccaggtcagcatgaac 900  
Qy 1004 gatgacctcttttctcagttctgtcatcatcaccacacacactgttctgacggt 1063  
Db 901 gatgacctcttttctcagttctgtcatcatcaccacacacactgttctcccaaa 960  
Qy 1064 tc 1065  
Db 961 tc 962

RESULT 7  
US-10-038-307-23  
; Sequence 23, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 23  
; LENGTH: 1056  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-038-307-23

Query Match 67.3%; Score 951.4; DB 7; Length 1056;  
Best Local Similarity 99.9%; Pred. No. 1.3e-125;  
Matches 952; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 102 ccattgcccacgcgagcgagagccctcgcatcggttccagtggtctctttggcca 161  
Db 11 ccattgcccacgcgagcgagagccctcgcatcggttccagtggtctctttggcca 70

```
QY 162 ctctgtgctcatctcgcgcgcaaggagggagcagcagggagggagggggtccagctgct 221
|||||
Db 71 ctctggcgcctcatctcgcgcgcaaggagggagcagcagggagggggtccagctgct 130
|||||
QY 222 acggcgatttgacctgacttacttcttcttgacaaatcaggaaagtgtgctcacactgga 281
|||||
Db 131 acggcgatttgacctgacttacttcttcttgacaaatcaggaaagtgtgctcacactgga 190
|||||
QY 282 atgaatctattacttcttggaacagttggtctacaaatcaccagccacagttgagaa 341
|||||
Db 191 atgaatctattacttcttggaacagttggtctacaaatcaccagccacagttgagaa 250
|||||
QY 342 tgtccttattgtttctccaccgaggaacaccttaattgaaactgacagaagacagag 401
|||||
Db 251 tgtccttattgtttctccaccgaggaacaccttaattgaaactgacagaagacagag 310
|||||
QY 402 acaaaatccgtcaaggcctagaagaactccagaagtcttgcagggaggagacacttaca 461
|||||
Db 311 acaaaatccgtcaaggcctagaagaactccagaagtcttgcagggaggagacacttaca 370
|||||
QY 462 tgcataaggatttgaaaggccagtgacagatttattgaaacagacaagggtaca 521
|||||
Db 371 tgcataaggatttgaaaggccagtgacagatttattgaaacagacaagggtaca 430
|||||
QY 522 ggacagccagcgtcatcttctgtgactggtgagaaactccatgaagatctcttttct 581
|||||
Db 431 ggacagccagcgtcatcttctgtgactggtgagaaactccatgaagatctcttttct 490
|||||
QY 582 attcagagaggagcctaataaggctcgcagatcttgtgcaattgtttactgtttggtg 641
|||||
Db 491 attcagagaggagcctaataaggctcgcagatcttgtgcaattgtttactgtttggtg 550
|||||
QY 642 tgaagaatttcaatgagacacagctgcccggattgcggacagtaaggatcatgtttc 701
|||||
Db 551 tgaagaatttcaatgagacacagctgcccggattgcggacagtaaggatcatgtttc 610
|||||
QY 702 ccgtgaatgacggctttcagcgtctgaaggcatcaccatccactcaattttgaaagtct 761
|||||
Db 611 ccgtgaatgacggctttcagcgtctgaaggcatcaccatccactcaattttgaaagtct 670
|||||
QY 762 gcatagaattctgacagtgaaacatccacatcatatgtgagaggagatctttcaagt 821
|||||
Db 671 gcatagaattctgacagtgaaacatccacatcatatgtgagaggagatctttcaagt 730
|||||
QY 822 tcgtgagaggaaacggcttcagcgtctgcaaggcatcaccatccactcaattttgaaagtct 881
|||||
Db 731 tcgtgagaggaaacggcttcagcgtctgcaaggcatcaccatccactcaattttgaaagtct 790
|||||
QY 882 agataaatgactcgttcacactcaatgagaagcccttttctgtggaagacacttatttac 941
|||||
Db 791 agataaatgactcgttcacactcaatgagaagcccttttctgtggaagacacttatttac 850
|||||
QY 942 tgtgtccagcgcctattcttaataaagggtgcatgaaagctgcactccaggtcagcatga 1001
|||||
Db 851 tgtgtccagcgcctattcttaataaagggtgcatgaaagctgcactccaggtcagcatga 910
|||||
QY 1002 acgatggcctctctttatctccagttctgtcatcatcaccacacacactgt 1054
|||||
Db 911 acgatggcctctctttatctccagttctgtcatcatcaccacacacactgt 963
|||||
```

```
RESULT 8
US-10-038-307-19
; Sequence 19, Application US/10038307
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
```

```
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 1713
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-038-307-19

Query Match 67.3%; Score 951.4; DB 7; Length 1713;
Best Local Similarity 99.9%; Pred. No. 1.3e-125;
Matches 952; Conservative 0; Mismatches .1; Indels 0; Gaps 0;

QY 102 ccatggccacggcgagcgagcgcctcgccatcggttccagttggctctctttggcca 161
|||||
Db 11 ccatggccacggcgagcgagcgcctcgccatcggttccagttggctctctttggcca 70
|||||
QY 162 ctctgtgtctcatctcgcgcgcaaggggagcagcagggagggatgggttccagctgct 221
|||||
Db 71 ctctgtgtctcatctcgcgcgcaaggggagcagcagggagggatgggggtccagctgct 130
|||||
QY 222 acggcgatttgacctgacttacttcttgacaaatcagaaagtgtgctcacactgga 281
|||||
Db 131 acggcgatttgacctgacttacttcttgacaaatcagaaagtgtgctcacactgga 190
|||||
QY 282 atgaatctattacttcttggaacagttggtctacaaatcaccagccacagttgagaa 341
|||||
Db 191 atgaatctattacttcttggaacagttggtctacaaatcaccagccacagttgagaa 250
|||||
QY 342 tgtccttattgtttctccaccgaggaacaccttaattgaaactgacagaagacagag 401
|||||
Db 251 tgtccttattgtttctccaccgaggaacaccttaattgaaactgacagaagacagag 310
|||||
QY 402 acaaaatccgtcaaggcctagaagaactccagaagtcttgcagggaggagacacttaca 461
|||||
Db 311 acaaaatccgtcaaggcctagaagaactccagaagtcttgcagggaggagacacttaca 370
|||||
QY 462 tgcataaggatttgaaaggccagtgacagatttattgaaacagacaagggtaca 521
|||||
Db 371 tgcataaggatttgaaaggccagtgacagatttattgaaacagacaagggtaca 430
|||||
QY 522 ggacagccagcgtcatcttctgtgactggtgagaaactccatgaagatctcttttct 581
|||||
Db 431 ggacagccagcgtcatcttctgtgactggtgagaaactccatgaagatctcttttct 490
|||||
QY 582 attcagagaggagcctaataaggctcgcagatcttgtgcaattgtttactgtttggtg 641
|||||
Db 491 attcagagaggagcctaataaggctcgcagatcttgtgcaattgtttactgtttggtg 550
|||||
QY 642 tgaagaatttcaatgagacacagctgcccggattgcggacagtaaggatcatgtttc 701
|||||
Db 551 tgaagaatttcaatgagacacagctgcccggattgcggacagtaaggatcatgtttc 610
|||||
QY 702 ccgtgaatgacggctttcagcgtctgaaggcatcaccatccactcaattttgaaagtct 761
|||||
Db 611 ccgtgaatgacggctttcagcgtctgaaggcatcaccatccactcaattttgaaagtct 670
|||||
QY 762 gcatagaattctgacagtgaaacatccacatcatatgtgagaggagatctttatttac 821
|||||
Db 671 gcatagaattctgacagtgaaacatccacatcatatgtgagaggagatctttcaagt 730
|||||
QY 822 tcgtgagaggaaacggcttcagcgtctgcaaggcatcaccatccactcaattttgaaagtct 881
|||||
Db 731 tcgtgagaggaaacggcttcagcgtctgcaaggcatcaccatccactcaattttgaaagtct 790
|||||
QY 882 agataaatgactcgttcacactcaatgagaagcccttttctgtggaagacacttatttac 941
|||||
Db 791 agataaatgactcgttcacactcaatgagaagcccttttctgtggaagacacttatttac 850
|||||
QY 942 tgtgtccagcgcctattcttaataaagggtgcatgaaagctgcactccaggtcagcatga 1001
|||||
Db 851 tgtgtccagcgcctattcttaataaagggtgcatgaaagctgcactccaggtcagcatga 910
|||||
```

Qy 1002 acgatggcctctcttttctccagttctgtctatccatccaccacacactgt 1054  
|||||  
Db 911 acgatggcctctcttttctccagttctgtctatccatccaccacacactgt 963

RESULT 9  
US-10-038-307-25  
; Sequence 25, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 25  
; LENGTH: 1008  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-038-307-25

Query Match 67.3%; Score 951; DB 7; Length 1008;  
Best Local Similarity 99.5%; Pred. No. 1.5e-125;  
Matches 954; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
  
Qy 102 ccattggccacggcgagcgagccctcggtcagtcggtccagtcggtctctttggcca 161  
|||||  
Db 11 ccattggccacggcgagcgagccctcggtcagtcggtccagtcggtctctttggcca 70  
  
Qy 162 ccttggtctcatctgcgcggcgcaagggggacgcaggagatgggggtccagctgct 221  
|||||  
Db 71 ccttggtctcatctgcgcggcgcaagggggacgcaggagatgggggtccagctgct 130  
  
Qy 222 acggcggtattgacctgtactctattcttgacaaatcaggaaagtgtgctgccaccctgga 281  
|||||  
Db 131 acggcggtattgacctgtactctattcttgacaaatcaggaaagtgtgctgccaccctgga 190  
  
Qy 282 atgaatctattactttgttggaacagtttggtccacaaatccatccacacagttgagaa 341  
|||||  
Db 191 atgaatctattactttgttggaacagtttggtccacaaatccatccacacagttgagaa 250  
  
Qy 342 tatctcttattgtttctccaccgcaggacaaaccttaataagaaactgaacagacagag 401  
|||||  
Db 251 tgctctttattgtttctccaccgcaggacaaaccttaataagaaactgaacagacagag 310  
  
Qy 402 aacaaatccgtcaaggcctagaagaactccagaagttctgcaggagagacacttaca 461  
|||||  
Db 311 aacaaatccgtcaaggcctagaagaactccagaagttctgcaggagagacacttaca 370  
  
Qy 462 tgcatgaaggatttgaaggccagtgagcagattattattatgaacacagacagggtaca 521  
|||||  
Db 371 tgcatgaaggatttgaaggccagtgagcagattattattatgaacacagacagggtaca 430  
  
Qy 522 gqacagccagctcatctgtttgactgatggagaactccataaagatctctttct 581  
|||||  
Db 431 ggacagccagctcatctgtttgactgatggagaactccataaagatctctttct 490  
  
Qy 582 attcagagaggaggcctaagttctcgagatcttggtgcaattgtttactgttggtg 641  
|||||  
Db 491 attcagagaggaggcctaagttctcgagatcttggtgcaattgtttactgttggtg 550  
  
Qy 642 tgaagatttcaatgagacacagctggcccgattgcgagacagtaaggatcatgtgttcc 701  
|||||  
Db 551 tgaagatttcaatgagacacagctggcccgattgcgagacagtaaggatcatgtgttcc 610  
  
Qy 702 ccgtgaatgacggcttccaggtcttcgaagcatcatccactcaattttgaaagaagtcct 761  
|||||  
Db 611 ccgtgaatgacggcttccaggtcttcgaagcatcatccactcaattttgaaagaagtcct 670

Qy 762 gcatcgaattctagcagctgaacctccaccatatgtgcagagagagtcatttcaagttg 821  
|||||  
Db 671 gcatcgaattctagcagctgaacctccaccatatgtgcagagagagtcatttcaagttg 730  
  
Qy 822 tcgtgagaggaaaacggcttcgcagacatgccgcgaacgctggagacgggtcctctcagcttca 881  
|||||  
Db 731 tcgtgagaggaaaacggcttcgcagacatgccgcgaacgctggagacgggtcctctcagcttca 790  
  
Qy 882 agatcaatgactcgttcacacacatcaatgagaagccctttctgtggaagacacttattac 941  
|||||  
Db 791 agatcaatgactcgttcacacacatcaatgagaagccctttctgtggaagatacttattac 850  
  
Qy 942 tgtgtccagcgctctattctaaagaagtggcatgaaagctgcactccaggtcagcatga 1001  
|||||  
Db 851 tgtgtccagcgctctattctaaagaagtggcatgaaagctgcactccaggtcagcatga 910  
  
Qy 1002 acgatggcctctcttttctccagttctgtcatcaccacacacactgttctgac 1060  
|||||  
Db 911 acgatggcctctcttttctccagttctgtcatcaccacacacacagtttggtc 969

RESULT 10  
US-10-038-307-9  
; Sequence 9, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 1650  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-038-307-9

Query Match 67.2%; Score 950.8; DB 7; Length 1650;  
Best Local Similarity 99.3%; Pred. No. 1.6e-125;  
Matches 955; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
Qy 104 atggccacggcgagcgagagccctcggtccagtcggtccagtcggtctctttggccact 163  
|||||  
Db 1 atggccacggcgagcgagagccctcggtccagtcggtccagtcggtctctttggccact 60  
  
Qy 164 ctggtgctcatctgcgcggcgcaagggggacgcaggagagatgggggtccagctgctac 223  
|||||  
Db 61 ctgggtgctcatctgcgcggcgcaagggggacgcaggagagatgggggtccagctgctac 120  
  
Qy 224 ggcgattggtgactgtactctattttggacaaatcagggaagtgtgctgccactggaat 283  
|||||  
Db 121 ggcgattggtgactgtactctattttggacaaatcagggaagtgtgctgccactggaat 180  
  
Qy 284 gaaatctattactttgtggaacagttgggtccaaaattcatcagccacacagttgaaatg 343  
|||||  
Db 181 gaaatctattactttgtggaacagttgggtccaaaattcatcagccacacagttgaaatg 240  
  
Qy 344 tctttattgtttctccaccgcaggagaaacacttaataatgaaactgacagaagacagaa 403  
|||||  
Db 241 tctttattgtttctccaccgcaggagaaacacttaataatgaaactgacagaagacagaa 300  
  
Qy 404 caaatccgtcaaggcctagaagaactccagaagttctccagggaggagacacttacatg 463  
|||||  
Db 301 caaatccgtcaaggcctagaagaactccagaagttctccagggaggagacacttacatg 360  
  
Qy 464 catgaaggatttgaaggccagtgagcagatttattatgaaaaacagacaagggtacagg 523  
|||||

Db 361 catgaaggatttgaaggccagtgagcagatttattatgaaacacagacagaggtacagg 420  
Qy 524 acagcagcgctacattgcttgaactgatgagaaacctcatgaagatctctttttctat 583  
Db 421 acagcagcgctacattgcttgaactgatgagaaacctcatgaagatctctttttctat 480  
Qy 584 tcagagagggaggttaataagtgctcagagatttggtgcaattgtttactgtgttggtg 643  
Db 481 tcagagagggaggttaataagtgctcagagatttggtgcaattgtttactgtgttggtg 540  
Qy 644 aaagatttcaatgagacacagctgcccgggattgagacagtaaggtatcatgttttccc 703  
Db 541 aaagatttcaatgagacacagctgcccgggattgagacagtaaggtatcatgttttccc 600  
Qy 704 gtgaatgacggctttcaggtctgcaaggcgcacatccactcaattttgaaagaagtctgc 763  
Db 601 gtgaatgacggctttcaggtctgcaaggcgcacatccactcaattttgaaagaagtctgc 660  
Qy 764 atcgaaattctagcagctgaaccatccaccatattgtgcagagagtcatttcaagtgttc 823  
Db 661 atcgaaattctagcagctgaaccatccaccatattgtgcagagagtcatttcaagtgttc 720  
Qy 824 gtgagagaaacggcttccgacatcccgcaacgtggagaggtctctgcagcttcaag 883  
Db 721 gtgagagaaacggcttccgacatcccgcaacgtggagaggtctctgcagcttcaag 780  
Qy 884 atcaatgactcgttcacactcaatgagaagcccttttctgtggaagatctatttactg 943  
Db 781 atcaatgactcgttcacactcaatgagaagcccttttctgtggaagatctatttactg 840  
Qy 944 tctcagcgcctattttaaagaagtgtgcatgaaagctgcactccaggtcagcatgaac 1003  
Db 841 tctcagcgcctattttaaagaagtgtgcatgaaagctgcactccaggtcagcatgaac 900  
Qy 1004 gatggcctcttttattctccagttctgtcatcactcaacacacacactcttctccaaa 1063  
Db 901 gatggcctcttttattctccagttctgtcatcactcaacacacacactcttctccaaa 960  
Qy 1064 tc 1065  
Db 961 tc 962

RESULT 11  
US-10-038-307-21  
; Sequence 21, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTHMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038, 307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 21  
; LENGTH: 1047  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-038-307-21

Query Match 64.7%; Score 914.4; DB 7; Length 1047;  
Best Local Similarity 97.1%; Pred. No. 2.2e-120;  
Matches 952; Conservative 0; Mismatches 1; Indels 27; Gaps 1;  
Qy 102 ccatggccacggcgagcggagagccctcgccatcggtccagttccagttcgtctctttggcca 161  
Db 11 ccatggccacggcgagcggagagccctcgccatcggtccagttcgtctctcttttggcca 70  
Qy 162 ctctggtgctcatctgcgcggcgcaagggggacgagggaggtgggggtcca----- 214

Db 71 ctctggtgctcatctgcgcggcgcaagggggacgcgagggatgggggtccaatggatt 130  
Qy 215 -----gctgctacgcgaggtttgacctgtacttcatcttttgaca 254  
Db 131 acaaggtgacgatgacaaggccctgctacgcgaggtttgacctgtacttcatcttttgaca 190  
Qy 255 aatcaggaagtgtgtgcaccactggaatgaaatctattacttttggaaacagttggctc 314  
Db 191 aatcaggaagtgtgtgcaccactggaatgaaatctattacttttggaaacagttggctc 250  
Qy 315 acaaatctacgcccacacagttgagaatgtccctttattgtttttctccaccaggaacaa 374  
Db 251 acaaatctacgcccacacagttgagaatgtccctttattgtttttctccaccaggaacaa 310  
Qy 375 ccttaataaactgacagagacagagaaacaaatccgtcaagcctcaggaactccaga 434  
Db 311 ccttaataaactgacagagacagagaaacaaatccgtcaagcctcaggaactccaga 370  
Qy 435 aagttctgccagagggagacacttacatgatgaaggtttgaaagggccagtgagcaga 494  
Db 371 aagttctgccagagggagacacttacatgatgaaggtttgaaagggccagtgagcaga 430  
Qy 495 ttattatgaaaaacagacaaagggtacagacagccagcgtcactcattgtttgactgatg 554  
Db 431 ttattatgaaaaacagacaaagggttacagggacagccagcgtcactcattgtttgactgatg 490  
Qy 555 gagaactccatgaagatctctttttctattcagagagggaggttaattaggtctcgagatc 614  
Db 491 gagaactccatgaagatctctttttctattcagagagggaggttaattaggtctcgagatc 550  
Qy 615 ttggtgcaattgtttactgtgttggttgaaagatttcaatgagacacacagctggcccgga 674  
Db 551 ttggtgcaattgtttactgtgttggttgaaagatttcaatgagacacacagctggcccgga 610  
Qy 675 ttgcgacagtaaggtacatgtgttcccggtgaatgacggttttcaggtctcgcaaggca 734  
Db 611 ttgcgacagtaaggtacatgtgttcccggtgaatgacggttttcaggtctcgcaaggca 670  
Qy 735 tcctccactcaattttgaaagtcctgcacgcaaaattctcagcgttcagcgtgaacatccacca 794  
Db 671 tcctccactcaattttgaaagtcctgcacgcaaaattctcagcgttcagcgtgaacatccacca 730  
Qy 795 tatgtgcagagagtcatttcaagtgtgtgcagagaaacggcttccgacatcccgca 854  
Db 731 tatgtgcagagagtcatttcaagtgtgtgcagagaaacggcttccgacatcccgca 790  
Qy 855 acgtggacaggtctctgcaggttcaagatcaatgactcgtgcacactcaatgagaagc 914  
Db 791 acgtggacaggtctctgcaggttcaagatcaatgactcgtgcacactcaatgagaagc 850  
Qy 915 cctttctgtggaagacacttatttactgtgtccagcgcctatctttaaagaagtggca 974  
Db 851 cctttctgtggaagatacttatttactgtgtccagcgcctatctttaaagaagtggca 910  
Qy 975 tgaagctgcactccaggtcagcatgaaacgatggcctctcttttctccaggttctgca 1034  
Db 911 tgaagctgcactccaggtcagcatgaaacgatggcctctcttttctccaggttctgca 970  
Qy 1035 tcatacacaccacacactgt 1054  
Db 971 tcatacacaccacacactgt 990

RESULT 12  
US-09-629-469A-11705  
; Sequence 11705, Application US/09629469A  
; GENERAL INFORMATION:  
; APPLICANT: OTA, TOSHIO  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: NISHIKAWA, TETSUO  
; APPLICANT: HAYASHI, KOJI  
; APPLICANT: SAITO, KAORU

APPLICANT: YAMAMOTO, JUNICHI  
APPLICANT: ISHII, SHIZUKO  
APPLICANT: SUGIYAMA, TOMOYASU  
APPLICANT: WAKAMATSU, AI  
APPLICANT: NAGAI, KEIICHI  
APPLICANT: OTSUKI, TETSUJI  
TITLE OF INVENTION: PRIMERS FOR SYNTHESIZING FULL-LENGTH cDNA AND THEIR USE  
FILE REFERENCE: 084335/0123  
CURRENT APPLICATION NUMBER: US/09/629,469A  
CURRENT FILING DATE: 2000-07-28  
PRIOR APPLICATION NUMBER: JP 1999-248036  
PRIOR FILING DATE: 1999-07-29  
PRIOR APPLICATION NUMBER: JP 1999-300253  
PRIOR FILING DATE: 1999-08-27  
PRIOR APPLICATION NUMBER: JP 2000-118776  
PRIOR FILING DATE: 2000-01-11  
PRIOR APPLICATION NUMBER: JP 2000-183767  
PRIOR FILING DATE: 2000-05-02  
PRIOR APPLICATION NUMBER: JP 2000-241899  
PRIOR FILING DATE: 2000-06-09  
PRIOR APPLICATION NUMBER: 60/159,590  
PRIOR FILING DATE: 1999-10-18  
PRIOR APPLICATION NUMBER: 60/183,322  
PRIOR FILING DATE: 2000-02-17  
NUMBER OF SEQ ID NOS: 19025  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 11705  
LENGTH: 1436  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (380)..(1033)  
US-09-629-469A-11705

Query Match 62.9%; Score 890; DB 5; Length 1436;  
Best Local Similarity 99.3%; Pred. No. 5.7e-117;  
Matches 904; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Qy 1 aggaccgcgaggagggccgcgagatggcgctccctgagggctcgtgaggtcgctgaggtcgcg 60  
Db 41 aggaccgcgaggagggccgcgagatggcgctccctgagggctcgtgaggtcgcgaggtcgcg 100  
Qy 61 agcctggcgaagagcgaccctgctctcccgggctgctcccgccatggccacggcgagcg 120  
Db 101 agcctggcgaagagcgaccctgctctcccgggctgctcccgccatggccacggcgagcg 160  
Qy 121 gagagccctcgcatcggtctccagtgctctcttggccactctggtgctcatctcgcg 180  
Db 161 gagagccctcgcatcggtctccagtgctctc-acggccactctggtgctcatctcgcg 219  
Qy 181 cgggcaaggggagcagggaggatgggggtccagcctgctacggcgagttgacctgta 240  
Db 220 cgggcaaggggagcagggaggatgggggtccagcctgctacggcgagttgacctgta 279  
Qy 241 ctctcatttggacaaatcaggaaagtgtctgcaccactggaatgaaatctattattgt 300  
Db 280 ctctcatttggacaaatcaggaaagtgtctgcaccactggaatgaaatctattattgt 339  
Qy 301 ggaacagttggctcacaattcatcagccacagttgagaaatgtcttattgtttctc 360  
Db 340 ggaacagttggctcacaattcatcagccacagttgagaaatgtcttattgtttctc 399  
Qy 361 caccgcgaggaacaaacttaataactgacagaagacagagaacaaatccgtcaaggcct 420  
Db 400 caccgcgaggaacaaacttaataactgacagaagacagagaacaaatccgtcaaggcct 459  
Qy 421 aagaagaactccagaagaagtcttccaggaggagagacacttacatgcatgaaggattgaaag 480  
Db 460 aagaagaactccagaagaagtcttccaggaggagagacacttacatgcatgaaggattgaaag 519  
Qy 481 ggcagtgagcagatttattgaaacacagacaagggttacaggacagccgctcatcat 540

Db 520 ggccagtgagcagatttattgaaacacagacaagggttacaggacagctagcgtcatcat 579  
Qy 541 tgccttgactgatggagaactccatgaagatctcttttctatttcagagagggaggtctaa 600  
Db 580 tgccttgactgatggagaactccatgaagatctcttttctatttcagagagggaggtctaa 639  
Qy 601 taggtctcgagatctctggaatgtttactgtgttggtggaagatttcaatgagac 660  
Db 640 taggtctcgagatctctggaatgtttactgtgttggtggaagatttcaatgagac 699  
Qy 661 acagctggcccgagattgcggacagtaagatcatgtgtttccctggaatgacgggtttca 720  
Db 700 acagctggcccgagattgcggacagtaagatcatgtgtttccctggaatgacgggtttca 759  
Qy 721 ggctctgcaaggcatcatccactcaattttgaagaagtcctgcacgaaattctagcagc 780  
Db 760 ggctctgcaaggcatcatccactcaattttgaagaagtcctgcacgaaattctagcagc 819  
Qy 781 tgaaccatccaccatattgtcaggagagtcatttcaagtgtcgtgagaggaacggctt 840  
Db 820 tgaaccatccaccatattgtcaggagagtcatttcaagtgtcgtgagaggaacggctt 879  
Qy 841 ccgacatgcccgcaacgtggacaggggtctctctgcagcttcaagatcaatgactcggtcac 900  
Db 880 ccgacatgcccgcaacgtggacaggggtctctctgcagcttcaagatcaatgactcggtcac 939  
Qy 901 actcaatgag 910  
Db 940 actcagtaag 949

## RESULT 13

US-10-038-307-11  
; Sequence 11, Application US/10038307  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 1623  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-038-307-11

Query Match 61.4%; Score 867.6; DB 7; Length 1623;  
Best Local Similarity 97.9%; Pred. No. 7.9e-114;  
Matches 879; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Qy 168 tgcctcatctgcggcggaaggagggagcagggagagatgggggtccagcctctacggcg 227  
Db 38 taccacaattgtctctgtccggggagcagcaggagagatgggggtccagcctctacggcg 97  
Qy 228 gattgacctgtactctatttggacaaatcaggaagtgtctgcaccactggaatgaaa 287  
Db 98 gattgacctgtactctatttggacaaatcaggaagtgtctgcaccactggaatgaaa 157  
Qy 288 tctattacttttggaacagttggctcacaatttcatcagccacagttgagaatgtcct 347  
Db 158 tctattacttttggaacagttggctcacaatttcatcagccacagttgagaatgtcct 217  
Qy 348 ttattgttttctccaccggaggaacaaaccttaataactgacagaagacagagaacaaa 407  
Db 218 ttattgttttctccaccggaggaacaaaccttaataactgacagaagacagagaacaaa 277

Qy 408 tccgtcaaggcctagaagaactccagaaagtctctgcccagagagagacacttaacatgatg 467  
 Db 278 tccgtcaaggcctagaagaactccagaaagtctctgcccagagagagacacttaacatgatg 337  
 Qy 468 aagatttgaaggccagtgagcagatttattatgaagaacagaggtacaggacag 527  
 Db 338 aagatttgaaggccagtgagcagatttattatgaagaacagaggtacaggacag 397  
 Qy 528 ccaggtcatcatgtcttgactgatgagaaactccatgaagatctcttttctattcag 587  
 Db 338 ccaggtcatcatgtcttgactgatgagaaactccatgaagatctcttttctattcag 457  
 Qy 588 agaggaggtcctaagttctgcagatcttggtgcaattgtttactgtgtgtgtgaaag 647  
 Db 458 agaggaggtcctaagttctgcagatcttggtgcaattgtttactgtgtgtgtgaaag 517  
 Qy 648 attcaatgagacacagctgcccggattcgagacagtaaggtatcatgtttccctgga 707  
 Db 518 attcaatgagacacagctgcccggattcgagacagtaaggtatcatgtttccctgga 577  
 Qy 708 atgacgggtttcaggtctgcaaggtcatctccactcaattttgaagaagtcctgcacag 767  
 Db 578 atgacgggtttcaggtctgcaaggtcatctccactcaattttgaagaagtcctgcacag 637  
 Qy 768 aaattctagcagctgaacatccaccatattgtgcagagagatcttcaagttgtcgtga 827  
 Db 638 aaattctagcagctgaacatccaccatattgtgcagagagatcttcaagttgtcgtga 697  
 Qy 828 gaggaaacggcttcgacatgcccgaacgtggacaggtctctgcagcttcaagatca 887  
 Db 698 gaggaaacggcttcgacatgcccgaacgtggacaggtctctgcagcttcaagatca 757  
 Qy 888 atgactcgtctacatcaatgaagaagcccttttctgtggaagacatttactgtctc 947  
 Db 758 atgactcgtctacatcaatgaagaagcccttttctgtggaagacatttactgtctc 817  
 Qy 948 cagcgcctatctaaagaagttgcatgaagctgcactccagctcagcatgaacagatg 1007  
 Db 818 cagcgcctatctaaagaagttgcatgaagctgcactccagctcagcatgaacagatg 877  
 Qy 1008 gctctcttttactcagttctgtctatcaatcaccacacacacactgttctgacggttc 1065  
 Db 878 gctctcttttactcagttctgtctatcaatcaccacacacacactgttctcacaatc 935

RESULT 14  
 PCT-US02-08253-186  
 ; Sequence 186, Application PC/YUS0208253  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Carson-Walter, Eleanor  
 ; APPLICANT: St. Croix, Brad  
 ; APPLICANT: Vogelstein, Bert  
 ; APPLICANT: Kinzler, Kenneth  
 ; TITLE OF INVENTION: ENDOHELIAL CELL EXPRESSION PATTERNS  
 ; FILE REFERENCE: 1107.00179  
 ; CURRENT APPLICATION NUMBER: PCT/US02/08253  
 ; PRIOR FILING DATE: 2002-04-10  
 ; PRIOR APPLICATION NUMBER: 60/282,850  
 ; PRIOR FILING DATE: 2001-04-11  
 ; PRIOR APPLICATION NUMBER: 60/308,829  
 ; PRIOR FILING DATE: 2001-08-01  
 ; NUMBER OF SEQ ID NOS: 359  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 186  
 ; LENGTH: 5220  
 ; TYPE: DNA  
 ; ORGANISM: Mus musculus  
 PCT-US02-08253-186  
 Query Match 60.4%; Score 853.4; DB 1; Length 5220;  
 Best Local Similarity 87.3%; Pred. No. 7.1e-112;  
 Matches 935; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy 127 cctcggcatcgcttccagtggtctctcttggccactctgttgctctcatctgcgcgggca 186  
 Db 291 cctgggtgcgggctgcgggactctgcgtgcactcgtgcgtgtgcgcgggca 350  
 Qy 187 aggggacgcagggagatgggggtccagcctgcgtacgcgggatttgacctgtacttcat 246  
 Db 351 cggggcccgcggagatgggggaccagcttgctacggagattgcaccttacttcat 410  
 Qy 247 ttgggcaaatcaaggagtgctgcaccactggaatgaatctatttcttgggaaca 306  
 Db 411 cctggcaaatcaaggagtgctgcaccactggaatgaatctatttcttgggagca 470  
 Qy 307 gttgggtccaaatcatcagccacagttgagaatgtcctttattgttttctccaccg 366  
 Db 471 gttgggtccaaatcatcagccacagttgagaatgtcctttattgttttctccaccg 530  
 Qy 367 aggaacaaccttaataaactgacagaagacagagaacaatacctcgaaggcctagaaga 426  
 Db 531 agggacaaccttaataaactgaggaaggaagacagatccgacaaggcctagaaga 590  
 Qy 427 actccagaagtctccagagagacacttcatgcatgaaggatttgaaggccag 486  
 Db 591 gctccagaagtctccagagaggaacattacatgcacgaaggattcgagggccag 650  
 Qy 487 tgacagatttattatgaacaacagagaaggttacagagaccagcgcgtcatctgttt 546  
 Db 651 tgacagatttattatgaacaacaggaatcagacagcgcgtcatctgcgttt 710  
 Qy 547 gactgatggagaactcatgaagatctctttctctacagagagggaggttaagatgc 606  
 Db 711 gacggatgggagctgcacagagacctctctctctacagagagggaggttaagatgc 770  
 Qy 607 tcgagatctgtgtcaattgtttactgtgtgtgtgaaagatttcaatgagacacagct 666  
 Db 771 ccgagacctgtgtcgtattgtttactggtgtgtgtgaaagatttcaatgaaactcagt 830  
 Qy 667 gggccgagattgcgagacaaaggaatcgtgttcccgtaagattacggttccaggtct 726  
 Db 831 ggcctcgatttcagacagtaaggaccacggttctctctgtaacacagcgttccaggtct 890  
 Qy 727 gcaaggcatctccactcaattttgaagaagctcctgcacatcgaaattctagcagctgaacc 786  
 Db 891 ccaaggcatctccactcaattttgaagaagctcctgcacatcgaaattctggtggaacc 950  
 Qy 787 atccacatctgcaggagagtgatttcaagtgtgtgagagaaacggcttccgaca 846  
 Db 951 atccacatctgcaggagagtgatttcaagtgtgtgagagaaacggcttccgaca 1010  
 Qy 847 tgcgcgaacgtggacaggttctctgcagcttcaagatcaatgactcgttcacactcaa 906  
 Db 1011 tgcgcgaacgtggacaggttctctgcagcttcaagatcaatgactcgttcacactcaa 1070  
 Qy 907 tgagaagccctttcttggaagacacttattactgtgtccagcgcgtatctataaaga 966  
 Db 1071 tgagaagccctttcttggaagacacttattactgtgtccagcgcgtatctataaaga 1130  
 Qy 967 agttggcatgaagctgcactccagtcagcatgaacatggcctctcttttactccag 1026  
 Db 1131 agttggcatgaagctgcactccagtcagcatgaacatggcctctctttactccag 1190  
 Qy 1027 ttctgtcatcatcacacacacactgttctgacggttccatctctggccatcgccctgct 1086  
 Db 1191 ttctgtcatcatcacacacacactgttctgacggttccatctctggccatcgccctgct 1250  
 Qy 1087 gatcctgttctcctagcctggtctctctctctgtgtgtgtgtgtgtgtgtgtgtgtgt 1146  
 Db 1251 ggtcctctctcctggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1310  
 Qy 1147 tctgattatcaaggaggttccctcccccctcctcagagagagtgaggagaaa 1197  
 Db 1311 agtgatcatcaaggaggttccctcccccctcctcagagagagtgaggagaaa 1361

[illegible]

us-09-970-076-1.rnpn

Fri Aug 9 10:56:52 2002

-----

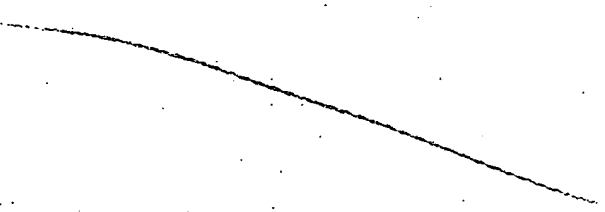


Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, .rnpm and .rnpn

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, .rapm and .rapn

***The Pending database search results should not be left in the case because they contain data that is confidential.***



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 8, 2002, 22:18:18 ; Search time 4997.37 Seconds  
(without alignments)  
6121.569 Million cell updates/sec

Title: US-09-970-076-1  
Perfect score: 1414  
Sequence: 1 aggaccgcgaggaaggcc.....aaaaaaaaaaaaaaaaaaaa 1414

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 21979536 seqs, 10817449327 residues  
Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_NA\_Main:  
1: /cgn2\_6/ptodata/2/pna/pctus\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/pna/US06\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/pna/US07\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/pna/US080\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/pna/US081\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/pna/US082\_COMB.seq.\*  
7: /cgn2\_6/ptodata/2/pna/US083\_COMB.seq.\*  
8: /cgn2\_6/ptodata/2/pna/US084\_COMB.seq.\*  
9: /cgn2\_6/ptodata/2/pna/US085\_COMB.seq.\*  
10: /cgn2\_6/ptodata/2/pna/US086\_COMB.seq.\*  
11: /cgn2\_6/ptodata/2/pna/US087\_COMB.seq.\*  
12: /cgn2\_6/ptodata/2/pna/US088\_COMB.seq.\*  
13: /cgn2\_6/ptodata/2/pna/US089\_COMB.seq.\*  
14: /cgn2\_6/ptodata/2/pna/US090\_COMB.seq.\*  
15: /cgn2\_6/ptodata/2/pna/US091\_COMB.seq.\*  
16: /cgn2\_6/ptodata/2/pna/US092\_COMB.seq.\*  
17: /cgn2\_6/ptodata/2/pna/US093\_COMB.seq.\*  
18: /cgn2\_6/ptodata/2/pna/US094\_COMB.seq.\*  
19: /cgn2\_6/ptodata/2/pna/US095A\_COMB.seq.\*  
20: /cgn2\_6/ptodata/2/pna/US095B\_COMB.seq.\*  
21: /cgn2\_6/ptodata/2/pna/US095C\_COMB.seq.\*  
22: /cgn2\_6/ptodata/2/pna/US095D\_COMB.seq.\*  
23: /cgn2\_6/ptodata/2/pna/US096A\_COMB.seq.\*  
24: /cgn2\_6/ptodata/2/pna/US096B\_COMB.seq.\*  
25: /cgn2\_6/ptodata/2/pna/US096C\_COMB.seq.\*  
26: /cgn2\_6/ptodata/2/pna/US096D\_COMB.seq.\*  
27: /cgn2\_6/ptodata/2/pna/US096E\_COMB.seq.\*  
28: /cgn2\_6/ptodata/2/pna/US097A\_COMB.seq.\*  
29: /cgn2\_6/ptodata/2/pna/US097B\_COMB.seq.\*  
30: /cgn2\_6/ptodata/2/pna/US097C\_COMB.seq.\*  
31: /cgn2\_6/ptodata/2/pna/US098A\_COMB.seq.\*  
32: /cgn2\_6/ptodata/2/pna/US098B\_COMB.seq.\*  
33: /cgn2\_6/ptodata/2/pna/US098C\_COMB.seq.\*  
34: /cgn2\_6/ptodata/2/pna/US099A\_COMB.seq.\*  
35: /cgn2\_6/ptodata/2/pna/US099B\_COMB.seq.\*  
36: /cgn2\_6/ptodata/2/pna/US099C\_COMB.seq.\*  
37: /cgn2\_6/ptodata/2/pna/US101\_COMB.seq.\*  
38: /cgn2\_6/ptodata/2/pna/US101\_COMB.seq.\*  
39: /cgn2\_6/ptodata/2/pna/US6000\_COMB.seq.\*  
40: /cgn2\_6/ptodata/2/pna/US6001\_COMB.seq.\*  
41: /cgn2\_6/ptodata/2/pna/US6002\_COMB.seq.\*  
42: /cgn2\_6/ptodata/2/pna/US6003\_COMB.seq.\*  
43: /cgn2\_6/ptodata/2/pna/US6004\_COMB.seq.\*

44: /cgn2\_6/ptodata/2/pna/US6005\_COMB.seq.\*  
45: /cgn2\_6/ptodata/2/pna/US6006\_COMB.seq.\*  
46: /cgn2\_6/ptodata/2/pna/US6007\_COMB.seq.\*  
47: /cgn2\_6/ptodata/2/pna/US6008\_COMB.seq.\*  
48: /cgn2\_6/ptodata/2/pna/US6009\_COMB.seq.\*  
49: /cgn2\_6/ptodata/2/pna/US6010\_COMB.seq.\*  
50: /cgn2\_6/ptodata/2/pna/US6011\_COMB.seq.\*  
51: /cgn2\_6/ptodata/2/pna/US6012\_COMB.seq.\*  
52: /cgn2\_6/ptodata/2/pna/US6013\_COMB.seq.\*  
53: /cgn2\_6/ptodata/2/pna/US6014\_COMB.seq.\*  
54: /cgn2\_6/ptodata/2/pna/US6015\_COMB.seq.\*  
55: /cgn2\_6/ptodata/2/pna/US6016\_COMB.seq.\*  
56: /cgn2\_6/ptodata/2/pna/US6017\_COMB.seq.\*  
57: /cgn2\_6/ptodata/2/pna/US6018\_COMB.seq.\*  
58: /cgn2\_6/ptodata/2/pna/US6019\_COMB.seq.\*  
59: /cgn2\_6/ptodata/2/pna/US6020\_COMB.seq.\*  
60: /cgn2\_6/ptodata/2/pna/US6021\_COMB.seq.\*  
61: /cgn2\_6/ptodata/2/pna/US6022\_COMB.seq.\*  
62: /cgn2\_6/ptodata/2/pna/US6023\_COMB.seq.\*  
63: /cgn2\_6/ptodata/2/pna/US6024\_COMB.seq.\*  
64: /cgn2\_6/ptodata/2/pna/US6025\_COMB.seq.\*  
65: /cgn2\_6/ptodata/2/pna/US6026\_COMB.seq.\*  
66: /cgn2\_6/ptodata/2/pna/US6027\_COMB.seq.\*  
67: /cgn2\_6/ptodata/2/pna/US6028\_COMB.seq.\*  
68: /cgn2\_6/ptodata/2/pna/US6029\_COMB.seq.\*  
69: /cgn2\_6/ptodata/2/pna/US6030\_COMB.seq.\*  
70: /cgn2\_6/ptodata/2/pna/US6031\_COMB.seq.\*  
71: /cgn2\_6/ptodata/2/pna/US6032\_COMB.seq.\*  
72: /cgn2\_6/ptodata/2/pna/US6033\_COMB.seq.\*  
73: /cgn2\_6/ptodata/2/pna/US6034\_COMB.seq.\*  
74: /cgn2\_6/ptodata/2/pna/US6035\_COMB.seq.\*  
75: /cgn2\_6/ptodata/2/pna/US6036\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1414	100.0	1414	36	US-09-970-076-1	Sequence 1, Appli
2	1193.8	84.4	2386	60	US-60-213-359-1222	Sequence 1222, Ap
3	1193.8	84.4	5540	34	US-09-918-715-176	Sequence 176, App
4	1193.8	84.4	5540	34	US-09-918-715-231	Sequence 231, App
5	1193.6	84.4	5540	36	US-09-970-076-5	Sequence 5, Appli
6	1192.4	84.3	2086	1	PCT-US00-30045-14	Sequence 14, Appl
7	1192.4	84.3	2086	1	PCT-US00-30045-45	Sequence 45, Appl
8	1052.4	74.4	2112	36	US-09-970-076-7	Sequence 7, Appli
9	1052.4	74.4	2272	1	PCT-US99-31025-8	Sequence 8, Appli
10	1052.4	74.4	2272	16	US-09-223-546-8	Sequence 8, Appli
11	1052.4	74.4	2272	18	US-09-471-179-8	Sequence 8, Appli
12	1052.4	74.4	2272	22	US-09-599-596-8	Sequence 8, Appli
13	1052.4	74.4	2272	30	US-09-796-753-11	Sequence 11, Appl
14	1052.4	74.4	2374	25	US-09-644-871-9264	Sequence 9264, Ap
15	1052.4	74.4	2374	28	US-09-710-281-4357	Sequence 4357, Ap
16	1052.4	74.4	2459	17	US-09-329-548A-68	Sequence 68, Appl
17	1052.4	74.4	2459	18	US-09-419-553-204	Sequence 204, App
18	1052.4	74.4	2459	19	US-09-516-448-998	Sequence 998, App
19	1052.4	74.4	2459	30	US-09-770-155-204	Sequence 204, App
20	1052.4	74.4	2459	30	US-09-780-183-68	Sequence 68, Appl
21	1052.4	74.4	2459	32	US-09-833-381-998	Sequence 998, App
22	1052.4	74.4	2459	32	US-09-833-382-998	Sequence 998, App
23	1050.8	74.3	2401	17	US-09-371-168-7727	Sequence 7727, Ap
24	1050.8	74.3	2401	25	US-09-644-868-7637	Sequence 7637, Ap
25	1050.8	74.3	2401	25	US-09-644-871-7309	Sequence 7309, Ap
26	1050.8	74.3	2401	25	US-09-649-163-6281	Sequence 6281, Ap
27	1050.8	74.3	2401	25	US-09-652-121-5593	Sequence 5593, Ap
28	1050.8	74.3	2401	25	US-09-652-127-7873	Sequence 7873, Ap
29	1050.8	74.3	2401	31	US-09-801-833-7727	Sequence 7727, Ap
30	1025.8	72.5	2485	26	US-09-668-337-188	Sequence 188, App
31	963	68.1	2397	24	US-09-625-102-1757	Sequence 1757, Ap

Fri Aug 9 10:56:52 2002

32 949.4 67.1 999 1 PCT-US99-31025-10 Sequence 10, Appl  
33 949.4 67.1 999 16 US-09-223-546-10 Sequence 10, Appl  
34 949.4 67.1 999 18 US-09-471-179-10 Sequence 10, Appl  
35 949.4 67.1 999 22 US-09-599-596-10 Sequence 9977, Ap  
36 906.8 64.1 1402 18 US-09-471-275-9977 Sequence 3907, Ap  
37 906.8 64.1 1402 18 US-09-488-725A-3907 Sequence 335, App  
38 906.8 64.1 1609 18 US-09-488-725A-335 Sequence 8, Appl  
39 906.8 64.1 1609 24 US-09-620-312B-8 Sequence 8, Appl  
40 906.8 64.1 1609 37 US-10-037-270-9 Sequence 186, App  
41 890 62.9 1436 36 US-09-970-076-9 Sequence 9, Appl  
42 853.4 60.4 5220 34 US-09-918-715-186 Sequence 300, App  
43 853.4 60.4 5220 34 US-09-918-715-300 Sequence 20845, A  
44 829.8 58.7 2981 75 US-60-360-207-20845 Sequence 176, App  
45 563.4 39.8 596 17 US-09-330-360-176

ALIGNMENTS

RESULT 1  
US-09-970-076-1  
; Sequence 1, Application US/09970076  
; GENERAL INFORMATION:  
; APPLICANT: Young, John A.T.  
; APPLICANT: Bradley, Kenneth A.  
; APPLICANT: Collier, Robert J.  
; APPLICANT: Mogridge, Jeremy S.  
; TITLE OF INVENTION: Anthrax Toxin Receptor  
; FILE REFERENCE: 960296.97745  
; CURRENT APPLICATION NUMBER: US/09/970,076  
; CURRENT FILING DATE: 2001-10-03  
; PRIOR APPLICATION NUMBER: 60/251,481  
; PRIOR FILING DATE: 2000-12-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; TYPE: DNA  
; LENGTH: 1414  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (104)..(1207)  
US-09-970-076-1

Query Match 100.0%; Score 1414; DB 36; Length 1414;  
Best Local Similarity 100.0%; Pred. No. le-156;  
Matches 1414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aggaccgcgaggaagcccgccgagatggcgctccctgagggctcgtgaggttcgcgg 60  
Db 1 aggaccgcgaggaagcccgccgagatggcgctccctgagggctcgtgaggttcgcgg 60  
QY 61 agcgtgggaagggagccctgctctcccccggctgcccagccatggccacggcgagcg 120  
Db 61 agcgtgggaagggagccctgctctcccccggctgcccagccatggccacggcgagcg 120  
QY 121 gagagccctggcatcggctccagtggtctcttttggcactctggtgctcatctgcgc 180  
Db 121 gagagccctggcatcggctccagtggtctcttttggcactctggtgctcatctgcgc 180  
QY 181 cgggcaagggggcgcgagggagatgggggtccagcctgctacggcgagattgacctgta 240  
Db 181 cgggcaagggggcgcgagggagatgggggtccagcctgctacggcgagattgacctgta 240  
QY 241 cttaattttggacaaataacgaagatgtgctgcaccactggaatgaaatctattacttgt 300  
Db 241 cttaattttggacaaataacgaagatgtgctgcaccactggaatgaaatctattacttgt 300  
QY 301 ggaacagtgtggctacaaattcatcagccacacagttgagaaatgtctttttctc 360  
Db 301 ggaacagtgtggctacaaattcatcagccacacagttgagaaatgtctttttctc 360

QY 361 caccgaggaacaacaccttaataactgaactgacagagacagagacaaataacogtcaagcct 420  
Db 361 caccgaggaacaacaccttaataactgaactgacagagacagagacaaataacogtcaagcct 420  
QY 421 agaagaactccagaagaattctgcaggaggagagacacttacaatgcatgaagatttgaag 480  
Db 421 agaagaactccagaagaattctgcaggaggagagacacttacaatgcatgaagatttgaag 480  
QY 481 ggcagtgagcagatattattatgaaacacagaaggggtacaggacagccagcgtcatcat 540  
Db 481 ggcagtgagcagatattattatgaaacacagaaggggtacaggacagccagcgtcatcat 540  
QY 541 tgccttgactgagagaaactccatgaagatctctttttattcagagagagggagctaa 600  
Db 541 tgccttgactgagagaaactccatgaagatctctttttattcagagagagggagctaa 600  
QY 601 taggtctcgagatcttggtgcaattgtttactgtgttggttgaaagatttcaatgagac 660  
Db 601 taggtctcgagatcttggtgcaattgtttactgtgttggttgaaagatttcaatgagac 660  
QY 661 acagctggccggattgcggacagtaagatcatgtgttcccggtgaatgacgggtttca 720  
Db 661 acagctggccggattgcggacagtaagatcatgtgttcccggtgaatgacgggtttca 720  
QY 721 ggctctgcaaggcatcatccatcaattttgaaagagtcctgcacgaaattctagcagc 780  
Db 721 ggctctgcaaggcatcatccatcaattttgaaagagtcctgcacgaaattctagcagc 780  
QY 781 tgaacatccaccatattgagagagagtcatttcaagttgctgagaggaacggctt 840  
Db 781 tgaacatccaccatattgagagagagtcatttcaagttgctgagaggaacggctt 840  
QY 841 ccgacatgcccgcacgctggacagggctcctgcagcttcaagatcaatgacctcgctac 900  
Db 841 ccgacatgcccgcacgctggacagggctcctgcagcttcaagatcaatgacctcgctac 900  
QY 901 actcaatgagaagcccttttctgtgaaagacacttatttactgtgtccagcgcctatt 960  
Db 901 actcaatgagaagcccttttctgtgaaagacacttatttactgtgtccagcgcctatt 960  
QY 961 aaaagaagtggcatgaaagctcaactccagggtcagcatgaaagatggcctctctttat 1020  
Db 961 aaaagaagtggcatgaaagctcaactccagggtcagcatgaaagatggcctctctttat 1020  
QY 1021 ctccagttctgcatcatcaccacacacactgttctgacggttccatcctggccatcgc 1080  
Db 1021 ctccagttctgcatcatcaccacacacactgttctgacggttccatcctggccatcgc 1080  
QY 1081 cctgctgactctgctcctgagccctggctcctcctctggttctggttctggccctctg 1140  
Db 1081 cctgctgactctgctcctgagccctggctcctcctctggttctggttctggccctctg 1140  
QY 1141 ctgcaactgtgattatcaagaggtccctccacccctgcccggagaggtgaggaataa 1200  
Db 1141 ctgcaactgtgattatcaagaggtccctccacccctgcccggagaggtgaggaataa 1200  
QY 1201 aataaaatacaagaagaagaagaataccacagaacacagataaacctaacacagc 1260  
Db 1201 aataaaatacaagaagaagaagaataccacagaacacagataaacctaacacagc 1260  
QY 1261 ccgtgcaacgtattttatcaaatgctctgaaaaatcaatgctcaatctagacagctttt 1320  
Db 1261 ccgtgcaacgtattttatcaaatgctctgaaaaatcaatgctcaatctagacagctttt 1320  
QY 1321 cctctagttccctgtattcaaatccaggtgtctaaacttcaataaataagctatatgaaat 1380  
Db 1321 cctctagttccctgtattcaaatccaggtgtctaaacttcaataaataagctatatgaaat 1380  
QY 1381 caaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1414  
Db 1381 caaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1414

```

RESULT      2
US-60-213-359-1222
; Sequence 1222, Application US/60213359
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, preeti
; APPLICANT: Dlep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polymor
; TITLE OF INVENTION: Identified Thereby
; FILE REFERENCE: GX-0015 P
; CURRENT APPLICATION NUMBER: US/60/213,359
; CURRENT FILING DATE: 2000-06-21
; NUMBER OF SEQ ID NOS: 7924
; SOFTWARE: PERL Program
; SEQ ID NO 1222
; LENGTH: 2386
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 243308.2
US-60-213-359-1222

```

```
Query Match      84.4%; Score 1193.8; DB 60; Length 2386;
Best Local Similarity 99.8%; Pred. No. 5.6e-131;
Matches 1195; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

[illegible]

```

RESULT      3
US-09-918-715-176
; Sequence 176, Application US/09918715
; GENERAL INFORMATION:
; APPLICANT: Brad St. Croix
; APPLICANT: Bert Vogelstein
; APPLICANT: Kenneth Kinzler
; TITLE OF INVENTION: ENDOETHLIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00134
; CURRENT APPLICATION NUMBER: US/09/918,715
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 60/222,599
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 60/224,360
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 176
; LENGTH: 3540
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-715-176

```

	Query Match	84.4%	Score 1193.8;	DB 34;	Length 5540;
	Best Local Similarity	99.8%	Pred. No. 5.1e-131;		
	Matches 1195;	Conservative	0;	Mismatches 2;	Indels 0; Gaps 0;
Qy	1	aggaccgcgaggaagggccgcgagatggcgctgccctgagggctgcgagtcgcgagttccgcgg	60		
Db	41	aggaccgcgaggaagggccgcgagatggcgctgccctgagggctgcgagtcgcgagttccgcgg	100		
Qy	61	agcgtgggaagagcgcgacctgtctctcccgggctgcggccatggccacggcggaagcgcg	120		

[illegible]



QY 1141 ctgactgtattatcaagaggtccctccacccctgcccagagagagtgaggaaa 1197  
 Db 1181 ctgactgtattatcaagaggtccctccacccctgcccagagagagtgaggaaa 1237

RESULT 6  
 PCT-US00-30045-14  
 ; Sequence 14, Application PC/TUS0030045  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Human Genome Sciences, Inc.  
 ; TITLE OF INVENTION: 28 Human Secreted Proteins  
 ; FILE REFERENCE: PS708PCT  
 ; CURRENT APPLICATION NUMBER: PCT/US00/30045  
 ; CURRENT FILING DATE: 2000-11-01  
 ; PRIOR APPLICATION NUMBER: 60/163,581  
 ; PRIOR FILING DATE: 1999-11-05  
 ; PRIOR APPLICATION NUMBER: 60/215,133  
 ; PRIOR FILING DATE: 2000-06-30  
 ; NUMBER OF SEQ ID NOS: 201  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 14  
 ; LENGTH: 2447  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; PCT-US00-30045-14

Query Match 84.4%; Score 1193.6; DB 1; Length 2447;  
 Best Local Similarity 99.7%; Pred. No. 5.9e-131;  
 Matches 1196; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 aggaaccgcgaggaagggccgcgagatggcgctccctgagggctgctgaggttcgcgagtcgcgg 60  
 Db 37 aggaaccgcgaggaagggccgcgagatggcgctccctgagggctgctgaggttcgcgagtcgcgg 96

QY 61 agcgtggaagagagagagagagagagagagagagagagagagagagagagagagagagagagag 120  
 Db 97 agcgtggaagagagagagagagagagagagagagagagagagagagagagagagagagagagag 156

QY 121 gagagccctgcgcagctggtccagtggtccagtggtccagtggtccagtggtccagtggtccagtggt 180  
 Db 157 gagagccctgcgcagctggtccagtggtccagtggtccagtggtccagtggtccagtggtccagtggt 216

QY 181 cggcgaagggagagagagagagagagagagagagagagagagagagagagagagagagagagagag 240  
 Db 217 cggcgaagggagagagagagagagagagagagagagagagagagagagagagagagagagagagag 276

QY 241 ctctcatttggacaaatcaggaagtgctgcacacactggaatgaaatcattacttgt 300  
 Db 277 ctctcatttggacaaatcaggaagtgctgcacacactggaatgaaatcattacttgt 336

QY 301 ggaacagttggtccacaaatcaggaagtggtccagtggtccagtggtccagtggtccagtggtccag 360  
 Db 337 ggaacagttggtccacaaatcaggaagtggtccagtggtccagtggtccagtggtccagtggtccag 396

QY 361 caccgagggagacaccttaataatgaaatcaggaagtggtccagtggtccagtggtccagtggtccag 420  
 Db 397 caccgagggagacaccttaataatgaaatcaggaagtggtccagtggtccagtggtccagtggtccag 456

QY 421 agaagaactccagaagttctgcagagagagagagagagagagagagagagagagagagagagagagag 480  
 Db 457 agaagaactccagaagttctgcagagagagagagagagagagagagagagagagagagagagagagag 516

QY 481 ggcagtgagcagatttattatgaaacagagagagagagagagagagagagagagagagagagagagag 540  
 Db 517 ggcagtgagcagatttattatgaaacagagagagagagagagagagagagagagagagagagagagag 576

QY 541 tgccttgactgagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 600  
 Db 577 tgccttgactgagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 636

QY 601 tagctctgagatcttggtgcaattgttactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 660

Db 637 taggtctcgagatottgtgtgcaattgttactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 696

QY 661 acagtggtcccgagattggtgagagagagagagagagagagagagagagagagagagagagagagagag 720

Db 697 acagtggtcccgagattggtgagagagagagagagagagagagagagagagagagagagagagagagag 756

QY 721 ggctctgcaagggcatcaccctcaattttgaagaagtcctgcatcgaatttctagcagc 780

Db 757 ggctctgcaagggcatcaccctcaattttgaagaagtcctgcatcgaatttctagcagc 816

QY 781 tgaacacatccaccatattgtgagagagagagagagagagagagagagagagagagagagagagagagag 840

Db 817 tgaacacatccaccatattgtgagagagagagagagagagagagagagagagagagagagagagagagag 876

QY 841 ccgacatgcccgcaacgtgagagagagagagagagagagagagagagagagagagagagagagagagag 900

Db 877 ccgacatgcccgcaacgtgagagagagagagagagagagagagagagagagagagagagagagagagag 936

QY 901 actcaatgagaagccctttctgtggaagagagagagagagagagagagagagagagagagagagagagag 960

Db 937 actcaatgagaagccctttctgtggaagagagagagagagagagagagagagagagagagagagagagag 996

QY 961 aaaagaagttggcatgaaagctgcactccaggtcagcatgaaacagatggcctctcttttat 1020

Db 997 aaaagaagttggcatgaaagctgcactccaggtcagcatgaaacagatggcctctcttttat 1056

QY 1021 ctccagttctgcatcatcaccacacacacactgttctgacggttccatcctggtccatcgc 1080

Db 1057 ctccagttctgcatcatcaccacacacacactgttctgacggttccatcctggtccatcgc 1116

QY 1081 cctgctgactcctgtctgctcctagccctgctcctcctctggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1140

Db 1117 cctgctgactcctgtctgctcctagccctgctcctcctctggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1176

QY 1141 ctgactgtgattatcaagagaggtccctccacccctgcccagagagagagagagagagagagagagagag 1200

Db 1177 ctgactgtgattatcaagagaggtccctccacccctgcccagagagagagagagagagagagagagagag 1236

RESULT 7  
 PCT-US00-30045-45  
 ; Sequence 45, Application PC/TUS0030045  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Human Genome Sciences, Inc.  
 ; TITLE OF INVENTION: 28 Human Secreted Proteins  
 ; FILE REFERENCE: PS708PCT  
 ; CURRENT APPLICATION NUMBER: PCT/US00/30045  
 ; CURRENT FILING DATE: 2000-11-01  
 ; PRIOR APPLICATION NUMBER: 60/163,581  
 ; PRIOR FILING DATE: 1999-11-05  
 ; PRIOR APPLICATION NUMBER: 60/215,133  
 ; PRIOR FILING DATE: 2000-06-30  
 ; NUMBER OF SEQ ID NOS: 201  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 45  
 ; LENGTH: 2086  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (10)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: SITE  
 ; LOCATION: (2070)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: SITE  
 ; LOCATION: (2075)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: SITE  
 ; LOCATION: (2079)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: SITE  
 ; LOCATION: (2083)



OTHER INFORMATION: n equals a.t.g. or c  
PCT-US00-30045-45

Query Match 84.3%; Score 1192.4; DB 1; Length 2086;  
Best Local Similarity 99.4%; Pred. No. 8.4e-131;  
Matches 1193; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 aggaccgaggaagggccgagatggcgagctccctgaggtcgtggcgagttccgcg 60  
Db 63 aggaccgaggaagggccgagatggcgagctccctgaggtcgtggcgagttccgcg 122

Qy 61 agcgtgggaagagcgacccctgctctcccggtcgtggcgagttccgagcgagcg 120  
Db 123 agcgtgggaagagcgacccctgctctcccggtcgtggcgagttccgagcgagcg 182

Qy 121 gagagccctcggaatcggtctccagtggtctctcttggccactctggtctcatctcg 180  
Db 183 gagagccctcggaatcggtctccagtggtctctcttggccactctggtctcatctcg 242

Qy 181 cgggcaaggggagcgagggagatgggtccagctgctacgctgctacgctgattgacctga 240  
Db 243 cgggcaaggggagcgagggagatgggtccagctgctacgctgctacgctgattgacctga 302

Qy 241 ctctatttggacaaatcaggaaagtgtgctgacacactggaatgaaatctattacttgt 300  
Db 303 ctctatttggacaaatcaggaaagtgtgctgacacactggaatgaaatctattacttgt 362

Qy 301 ggaacagtgtgctcacaattcatcagccacagttgagaaatgctcttattgtttctc 360  
Db 363 ggaacagtgtgctcacaattcatcagccacagttgagaaatgctcttattgtttctc 422

Qy 361 caccggaggaacacctaataagaactgacagaagacagagaacaaatccgtcaagcct 420  
Db 423 caccggaggaacacctaataagaactgacagaagacagagaacaaatccgtcaagcct 482

Qy 421 agaagaactccagaaagtctgacagagagagacacttacatgcatgaagatttgaag 480  
Db 483 agaagaactccagaaagtctgacagagagagacacttacatgcatgaagatttgaag 542

Qy 481 ggcagtgagcagatttattatgaaacagacaaaggtacagagcagccagcgtcatcat 540  
Db 543 ggcagtgagcagatttattatgaaacagacaaaggtacagagcagccagcgtcatcat 602

Qy 541 tgccttactgagagaaactccatgaatctcttcttcttcttcttcttcttcttcttct 600  
Db 603 tgccttactgagagaaactccatgaatctcttcttcttcttcttcttcttcttcttct 662

Qy 601 taggtctcgagatcttggtgcaattgtttactgtgtgtgtgaaagatttcaatgagac 660  
Db 663 taggtctcgagatcttggtgcaattgtttactgtgtgtgtgaaagatttcaatgagac 722

Qy 661 acagctggcccgagatttcgagcagtaagatcatgtgttcccgtaagacgctttca 720  
Db 723 acagctggcccgagatttcgagcagtaagatcatgtgttcccgtaagacgctttca 782

Qy 721 ggcctctgaaagcagatccactcaatttgaagaagcctgcatcgaaatcttagcagc 780  
Db 783 ggcctctgaaagcagatccactcaatttgaagaagcctgcatcgaaatcttagcagc 842

Qy 781 tgaaccatccaccatattgtgagagagatcatttcaagttgtctgagagaaacgcttt 840  
Db 843 tgaaccatccaccatattgtgagagagatcatttcaagttgtctgagagaaacgcttt 902

Qy 841 ccgacatcccgcaacgtggacagggctctctgagcttcaagatcaatgactcggtcac 900  
Db 903 ccgacatcccgcaacgtggacagggctctctgagcttcaagatcaatgactcggtcac 962

Qy 901 actcaatgagaagccctttctgtggaagacacttatttactgtgtcagcgctattctt 960  
Db 963 actcaatgagaagccctttctgtggaagacacttatttactgtgtcagcgctattctt 1022

Qy 961 aaaagaagtggcatgaaagtgcactccaggtgcagcatgaacgatggcctctctttat 1020

Db 1023 aaaagaagtggcatgaaagtgcactccaggtcagcatgaacgatgacctctttat 1082

Qy 1021 ctccagttctgcatcatcaccaccacacactgttctgaaggttccatctcctgacatcgc 1080

Db 1083 ctccagttctgcatcatcaccaccacacactgttctgaaggttccatctcctgacatcgc 1142

Qy 1081 cctgctgacatctgtctctgctcctagcctggtctctctctctggttctggtccctctg 1140

Db 1143 cctgctgacatctgtctctgctcctagcctggtctctctctggttctggtccctctg 1202

Qy 1141 ctgcaactgtgattatcaagaggttccctccaccctcccgagagagagtgaggaaaaaa 1200

Db 1203 ctgcaactgtgattatcaagaggttccctccaccctcccgagagagagtgaggtaag 1262

RESULT 8  
US-09-970-076-7  
; Sequence 7, Application US/09970076  
; GENERAL INFORMATION:  
; APPLICANT: Young, John A.T.  
; APPLICANT: Bradley, Kenneth A.  
; APPLICANT: Collier, Robert J.  
; APPLICANT: Mogridge, Jeremy S.  
; TITLE OF INVENTION: Anthrax Toxin Receptor  
; FILE REFERENCE: 960296.97745  
; CURRENT APPLICATION NUMBER: US/09/970.076  
; CURRENT FILING DATE: 2001-10-03  
; PRIOR APPLICATION NUMBER: 60/251,481  
; PRIOR FILING DATE: 2000-12-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 2112  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (113)..(1111)  
US-09-970-076-7

Query Match 74.4%; Score 1052.4; DB 36; Length 2112;  
Best Local Similarity 99.9%; Pred. No. 2e-114;  
Matches 1053; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 aggaccgaggaagggcccgagatggcgagctccctgaggtcgtggcgagttccgcg 60  
Db 10 aggaccgaggaagggcccgagatggcgagctccctgaggtcgtggcgagttccgcg 69

Qy 61 agcgtgggaagagcgagccctgctctcccggtcgtggcgagttccgagcgagcg 120  
Db 70 agcgtgggaagagcgagccctgctctcccggtcgtggcgagttccgagcgagcg 129

Qy 121 gagagccctcggaatcggtctccagtggtctcttggccactctggtgctcatctcg 180  
Db 130 gagagccctcggaatcggtctccagtggtctcttggccactctggtgctcatctcg 189

Qy 181 cgggcaaggggagcgagggagatgggggtccagcctgctacgctgattgacctga 240  
Db 190 cgggcaaggggagcgagggagatgggggtccagcctgctacgctgattgacctga 249

Qy 241 ctctatttggacaaatcaggaaagtgtgctgacacactggaatgaaatctattacttgt 300  
Db 250 ctctatttggacaaatcaggaaagtgtgctgacacactggaatgaaatctattacttgt 309

Qy 301 ggaacagtgtgctcacaattcatcagccacagttcagaaatgctcttattttctc 360  
Db 310 ggaacagtgtgctcacaattcatcagccacagttcagaaatgctcttattttctc 369

Qy 361 caccggaggaacacctaataagaactgacagaagacagagaacaaatccgtcaagcct 420  
Db 370 caccggaggaacacctaataagaactgacagaagacagagaacaaatccgtcaagcct 429

*Handwritten signature*



```
RESULT 10
US-09-223-546-8
; Sequence 8, Application US/09223546
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 09404/056001
; CURRENT APPLICATION NUMBER: US/09/223,546
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 2272
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (213)...(1211)
US-09-223-546-8

Query Match          74.4%; Score 1052.4; DB 16; Length 2272;
Best Local Similarity 99.9%; Pred. No. 2e-114;
Matches 1053; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 aggccgcgaggaagggccgcgagatggcgctccctcctgaggtcgtggcgagttcgcg 60
Db 110 agggccgcgcgaggaagggccgcgagatggcgctccctcctgaggtcgtggcgagttcgcg 169
QY 61 agcgtgggaagagcgaccctgctctcccggtcgtcccgccatggccacggcgagcg 120
Db 170 agcgtgggaagagcgaccctgctctcccggtcgtcccgccatggccacggcgagcg 229
QY 121 gagagccctcgacatggcctccagtggtctctttggccactcgttgctcattctgcgc 180
Db 230 gagagccctcgacatggcctccagtggtctctttggccactcgttgctcattctgcgc 289
QY 181 cgggcaagggggagcgagggagggatgggggtccagcctcgtcacggcggtattggaacctgta 240
Db 290 cgggcaagggggagcgagggagggatgggggtccagcctcgtcacggcggtattggaacctgta 349
QY 241 cttcattttggacaaaatcagggaagtgtctgcaccactgggaatgaaatctattactttgt 300
Db 350 cttcattttggacaaaatcagggaagtgtctgcaccactgggaatgaaatctattactttgt 409
```

```
QY 721 ggctctgaaggcatcatccactcaattttgaagaagtcctcgtcatcgaaaaattctagcgc 780
Db 830 ggctctgaaggcatcatccactcaattttgaagaagtcctcgtcatcgaaaaattctagcgc 889
QY 781 tgaaccatccaccatattgtaggagagtcatttcaagttgtcgtgagagaaacggtt 840
Db 890 tgaaccatccaccatattgtaggagagtcatttcaagttgtcgtgagagaaacggtt 949
QY 841 ccgacatgcccgcaacgtagacaggggtcctcgtcagettcaagatcaatgactcgggtcac 900
Db 950 ccgacatgcccgcaacgtagacaggggtcctcgtcagettcaagatcaatgactcgggtcac 1009
QY 901 actcaatgagaagccctttcttctgtggaagacattatttactgtgtccagcgctattctt 960
Db 1010 actcaatgagaagccctttcttctgtggaagacattatttactgtgtccagcgctattctt 1069
QY 961 aaaagaagttggcatgaaagctgcactccaggtcagcatgaacgatggcctctcttttat 1020
Db 1070 aaaagaagttggcatgaaagctgcactccaggtcagcatgaacgatggcctctcttttat 1129
QY 1021 ctcagttctgtcatcatcaccacacacactgt 1054
Db 1130 ctcagttctgtcatcatcaccacacacactgt 1163

RESULT 11
US-09-471-179-8
; Sequence 8, Application US/09471179
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 7853-173
; CURRENT APPLICATION NUMBER: US/09/471,179
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 2272
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (213)...(1211)
US-09-471-179-8

Query Match          74.4%; Score 1052.4; DB 18; Length 2272;
Best Local Similarity 99.9%; Pred. No. 2e-114;
Matches 1053; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 aggccgcgaggaagggccgcgagatggcgctccctcctgaggtcgtggcgagttcgcg 60
Db 110 aggccgcgaggaagggccgcgagatggcgctccctcctgaggtcgtggcgagttcgcg 169
QY 61 agcgtgggaagagcgaccctgctctcccggtcgtcccgccatggccacggcgagcg 120
Db 170 agcgtgggaagagcgaccctgctctcccggtcgtcccgccatggccacggcgagcg 229
QY 121 gagagccctcgacatggcctccagtggtctctttggccactcgttgctcattctgcgc 180
Db 230 gagagccctcgacatggcctccagtggtctctttggccactcgttgctcattctgcgc 289
QY 181 cgggcaagggggagcgagggagggatgggggtccagcctcgtcacggcggtattggaacctgta 240
Db 290 cgggcaagggggagcgagggagggatgggggtccagcctcgtcacggcggtattggaacctgta 349
QY 241 cttcattttggacaaaatcagggaagtgtctgcaccactgggaatgaaatctattactttgt 300
Db 350 cttcattttggacaaaatcagggaagtgtctgcaccactgggaatgaaatctattactttgt 409
```

Fri Aug 9 10:56:52 2002

us-09-970-076-1.rnmp

QY 301 ggaacagttggtcacaattcattcaatgagccacagttgagaatgtcctttattgtttttc 360  
Db 410 ggaacagttggtcacaattcattcaatgagccacagttgagaatgtcctttattgtttttc 469  
QY 361 caccgagggaacaaccttaataatgaaactgacagaagacagagaacaataatcgtcaaggcct 420  
Db 470 caccgagggaacaaccttaataatgaaactgacagaagacagagaacaataatcgtcaaggcct 529  
QY 421 agaagaactcagaagaagttctgcagagagagacacttacatcatgaaggatttgaag 480  
Db 530 agaagaactcagaagaagttctgcagagagagacacttacatcatgaaggatttgaag 589  
QY 481 ggcagtgagcagattttatttataaacaagacagaaggtacagagaagcagcagcgtcat 540  
Db 590 ggcagtgagcagattttatttataaacaagacagaaggtacagagaagcagcagcgtcat 649  
QY 541 tgccttgactgagagaactccatgaagatctctttttctattcagagaagggaggtctaa 600  
Db 650 tgccttgactgagagaactccatgaagatctctttttctattcagagaagggaggtctaa 709  
QY 601 tagtctcgagatcttggtgcaattgtttactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 660  
Db 710 tagtctcgagatcttggtgcaattgtttactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 769  
QY 661 acagctggccggattgaggagatgaaggatcatgtgtttcccgatgaacggtttca 720  
Db 770 acagctggccggattgaggagatgaaggatcatgtgtttcccgatgaacggtttca 829  
QY 721 ggcctgcaaggcatcacccactcaatttgaagaagtcctgcacgcaaaattcttagcagc 780  
Db 830 ggcctgcaaggcatcacccactcaatttgaagaagtcctgcacgcaaaattcttagcagc 889  
QY 781 tgaaccatccaccatcatgtgcagagagatcatttcaagttgtgtgtgtgtgtgtgtgtgt 840  
Db 890 tgaaccatccaccatcatgtgcagagagatcatttcaagttgtgtgtgtgtgtgtgtgtgt 949  
QY 841 ccgacatgccgcaacgtggaagagtcctctgcagctcgaatcgaatgactcgggtcac 900  
Db 950 ccgacatgccgcaacgtggaagagtcctctgcagctcgaatcgaatgactcgggtcac 1009  
QY 901 actcaatgagaagcctttctgtggaagacacttatttactgtgtcagcgcctatctt 960  
Db 1010 actcaatgagaagcctttctgtggaagacacttatttactgtgtcagcgcctatctt 1069  
QY 1021 ctccagttctgtcatcatcaccaccacacactgt 1054  
Db 1130 ctccagttctgtcatcatcaccaccacacactgt 1163

RESULT 12  
US-09-599-596-8  
; Sequence 8, Application US/09599596  
; GENERAL INFORMATION:  
; APPLICANT: Holtzman, Douglas  
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM  
; FILE REFERENCE: 09404/066001  
; CURRENT APPLICATION NUMBER: US/09/599,596  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 09/223,546  
; PRIOR FILING DATE: 1998-12-30  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 8  
; LENGTH: 2272  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (213)...(1211)

US-09-599-596-8  
Query Match 74.4%; Score 1052.4; DB 22; Length 2272;  
Best Local Similarity 99.9%; Pred. No. 2e-114;  
Matches 1053; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 aggaccgcgaggaagggcccgagtgagcgtcccttgagggtcgtggcgagttcgcg 60  
Db 110 aggaccgcgaggaagggcccgagtgagcgtcccttgagggtcgtggcgagttcgcg 169  
QY 61 agcgtgggaagagcgaccctgtctcccgggctggggccatgccagcgcgagcg 120  
Db 170 agcgtgggaagagcgaccctgtctcccgggctggggccatgccagcgcgagcg 229  
QY 121 gagaacctcgactcgcttccagtggtctctttggccactctgggtcctcatctgc 180  
Db 230 gagaacctcgactcgcttccagtggtctctttggccactctgggtcctcatctgc 289  
QY 181 cgggcaaggggacgcagggaggaatgggggtccagcctgctacggcgatttgacctga 240  
Db 290 cgggcaaggggacgcagggaggaatgggggtccagcctgctacggcgatttgacctga 349  
QY 241 cttcattttggacaataatcaggaagtgtgtccacacaggaaatgaattctatttct 300  
Db 350 cttcattttggacaataatcaggaagtgtgtccacacaggaaatgaattctatttct 409  
QY 301 ggaacagttggtcacaataatcatcagccacagttgagaatgtcctttattgtttctc 360  
Db 410 ggaacagttggtcacaataatcatcagccacagttgagaatgtcctttattgtttctc 469  
QY 361 caccgaggaacaaccttaataatgaaactgacagaagacagacaataatcgtcaaggcct 420  
Db 470 caccgaggaacaaccttaataatgaaactgacagaagacagacaataatcgtcaaggcct 529  
QY 421 agagaactccagaagttctgcagagagagacacttacatgcatgaaggatttgaag 480  
Db 530 agagaactccagaagttctgcagagagagacacttacatgcatgaaggatttgaag 589  
QY 481 ggcagtgagcagattttatttataaacaagacagaaggtacagacagcagcagcgtcat 540  
Db 590 ggcagtgagcagattttatttataaacaagacagaaggtacagacagcagcagcgtcat 649  
QY 541 tgccttgactgagagaactccatgaagatctctttttctattcagagaagggaggtctaa 600  
Db 650 tgccttgactgagagaactccatgaagatctctttttctattcagagaagggaggtctaa 709  
QY 601 tagtctcgagatcttggtgcaattgtttactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 660  
Db 710 tagtctcgagatcttggtgcaattgtttactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 769  
QY 661 acagctggccggattgagacagtaaggatcatgtgtttcccgatgaatgacggtttca 720  
Db 770 acagctggccggattgagacagtaaggatcatgtgtttcccgatgaatgacggtttca 829  
QY 721 ggcctgcaaggcatcacccactcaatttgaagaagtcctgcacgcaaaattcttagcagc 780  
Db 830 ggcctgcaaggcatcacccactcaatttgaagaagtcctgcacgcaaaattcttagcagc 889  
QY 781 tgaaccatccaccatcatgtgcagagagatcatttcaagttgtgtgtgtgtgtgtgtgtgt 840  
Db 890 tgaaccatccaccatcatgtgcagagagatcatttcaagttgtgtgtgtgtgtgtgtgtgt 949  
QY 841 ccgacatgccgcaacgtggaagagtcctctgcagctcgaatcgaatgactcgggtcac 900  
Db 950 ccgacatgccgcaacgtggaagagtcctctgcagctcgaatcgaatgactcgggtcac 1009  
QY 901 actcaatgagaagcctttctgtggaagacacttatttactgtgtcagcgcctatctt 960  
Db 1010 actcaatgagaagcctttctgtggaagacacttatttactgtgtcagcgcctatctt 1069  
QY 961 aaaagaagttgcatgaagcgtcagctccaggtcagcagatgaacgatgacctcttttat 1020



```

QY   961 aaaagaattggcatgaagctgcaactcaggctcagcatgaacgatggcctctctttat 1020
      |||
Db   1070 aaagaagtggcatgaagctgcaactcaggctcagcatgaacgatggcctctctttat 1129
      |||

QY   1021 ctccagttctgtcatcatcaccaccacacctgt 1054
      |||
Db   1130 ctccagttctgtcatcatcaccaccacacctgt 1163
      |||

RESULT 14
US-09-644-871-9264
; Sequence 9264, Application US/09644871
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: 1600.1167-001
; CURRENT APPLICATION NUMBER: US/09/644, 871
; CURRENT FILING DATE: 2000-08-28
; PRIOR APPLICATION NUMBER: 60/151,059
; PRIOR FILING DATE: 1999-08-27
; NUMBER OF SEQ ID NOS: 9739
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9264
; LENGTH: 2374
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-644-871-9264

Query Match          74.4%; Score 1052.4; DB 25; Length 2374;
Best Local Similarity 99.9%; Pred. No. 2e-114;
Matches 1053; Conservative 0; Mismatches 1; Indels 0; Gaps

QY   1 aggaccgcgaaagaagggcccgcgcatggcgctccctcagggtcgtgaggttcgcg 60
      |||
Db   223 adgaccgcgaaagaagggcccgcgcatggcgctccctcagggtcgtgaggttcgcg 282
      |||

QY   61 agcgtgggaagagcggaccctgctctccccgggtgcccagtgccacggcgagagcg 120
      |||
Db   283 agcgtgggaagagcggaccctgctctccccgggtgcccagtgccacggcgagagcg 342
      |||

QY   121 gagagccctcggcatcgggttccagtgctctctttggccactcgtgtcctatctgcgc 180
      |||
Db   343 gagagccctcggcatcgggttccagtgctctctttggccactcgtgtcctatctgcgc 402
      |||

QY   181 cggccaaggggacgcagaggagtagtgggtccagcctgctacggcgagattgaacctgta 240
      |||
Db   403 cgggcaagggggcgacgagggagtagtgggtccagcctgctacggcgagattgaacctgta 462
      |||

QY   241 ctctcattttggacaatacaggaaagtgtgtgcaccaactggaaatcatactatttgt 300
      |||
Db   463 ctctcattttggacaatacaggaaagtgtgtgcaccaactggaaatcatactatttgt 522
      |||

QY   301 ggaacagttggctcacaaattcatcagccacagtttgagaatgtcctttattgtttctc 360
      |||
Db   523 ggaacagttggctcacaaattcatcagccacagtttgagaatgtcctttattgtttctc 582
      |||

QY   361 caccggaggaaacaccttaataagactgcagaaagacagaaacaaatccgtcaaagcct 420
      |||
Db   583 caccggaggaaacaccttaataagactgcagaaagacagaaacaaatccgtcaaagcct 642
      |||

QY   421 agaagaactccagaagaagtgtgccaggagagacacttacatgatgaagattgaaag 480
      |||
Db   643 agaagaactccagaagaagtgtgccaggagagacacttacatgatgaagattgaaag 702
      |||

QY   481 gccacgtgagcagatttatatgaaaacagacaagggttacaggacagccagcgtcatcat 540
      |||
Db   703 gccacgtgagcagatttatatgaaaacagacaagggttacaggacagccagcgtcatcat 762
      |||

QY   541 tgccttgactgatggaaactccaatgaagatctcttttctatttcagagaggagggttaa 600

```

Search completed: August 9, 2002, 01:53:53  
Job time: 12935 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
 Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:46:19 ; Search time 102.68 seconds  
 (without alignments)  
 497.015 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_27\_321  
 Perfect score: 1536  
 Sequence: 1 GGGRRDGGPACYGFDLY.....GLSFISSSVIIITHCSDGS 295

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 17299429 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : SPTREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1536	100.0	368	Q96P02	Q96P02 homo sapien
2	574	37.4	245	Q96NC7	Q96nc7 homo sapien
3	388	25.3	97	Q96EC6	Q96ec6 homo sapien
4	159.5	10.4	1161	Q96YE7	Q96ye7 rattus norv
5	153.5	10.0	1332	Q96P08	Q96pq8 halocynthia
6	139	9.0	724	Q04588	Q04588 eimeria max
7	134.5	8.8	712	Q43981	Q43981 eimeria ten
8	134	8.7	760	Q70350	Q70350 mus musculu
9	132.5	8.6	765	Q98J99	Q98j9 neospora ca
10	131.5	8.6	517	Q43853	Q43853 homo sapien
11	131	8.5	660	Q9UGC3	Q9ugc3 homo sapien
12	130.5	8.5	920	Q28984	Q28984 sus scrofa
13	128	8.3	221	Q99L17	Q99l17 mus musculu
14	128	8.3	3767	Q9UAL3	Q9ual3 caenorhabdi
15	123	8.0	1151	Q9J130	Q9j130 rattus norv
16	118.5	7.7	3567	Q9ES77	Q9es77 mus musculu

17	117.5	7.6	1140	4	Q9P218	Q9p218 homo sapien
18	117.5	7.6	1207	4	Q9BOU7	Q9bqu7 homo sapien
19	117.5	7.6	2944	11	Q63870	Q63870 mus musculu
20	114.5	7.5	2104	5	Q21281	Q21281 caenorhabdi
21	114.5	7.5	2104	5	Q964N4	Q964n4 caenorhabdi
22	113.5	7.4	755	4	Q00261	Q00261 homo sapien
23	113.5	7.4	764	4	Q96HX6	Q96hx6 homo sapien
24	113.5	7.4	937	4	Q96FT5	Q96ft5 homo sapien
25	110	7.2	833	13	Q9YIC5	Q9yic5 cyprinus ca
26	109.5	7.1	704	10	Q9FF49	Q9ff49 arabidopsis
27	109.5	7.1	1905	5	Q9XTP6	Q9xtp6 plasmodium
28	109	7.1	1253	6	Q97566	Q97566 canis famil
29	108.5	7.1	956	11	Q99K64	Q99k64 mus musculu
30	108	7.0	650	11	Q9CV21	Q9cv21 mus musculu
31	106.5	6.9	562	5	Q01510	Q01510 plasmodium
32	106	6.9	415	11	Q923K3	Q923k3 mus musculu
33	106	6.9	833	5	Q96442	Q96442 strongyloce
34	105.5	6.9	567	5	Q18048	Q18048 caenorhabdi
35	104.5	6.8	769	5	Q00816	Q00816 toxoplasma
36	104.5	6.8	1169	11	Q9QXH4	Q9qxh4 mus musculu
37	104	6.8	205	11	Q63001	Q63001 rattus norv
38	104	6.8	747	13	Q91900	Q91900 xenopus lae
39	102	6.6	614	5	Q94674	Q94674 plasmodium
40	102	6.6	714	10	Q948G7	Q948g7 oryza sativ
41	102	6.6	759	4	Q96C67	Q96c67 homo sapien
42	101.5	6.6	425	5	Q9GZF5	Q9gzf5 caenorhabdi
43	101.5	6.6	537	4	Q96AA0	Q96aa0 homo sapien
44	101.5	6.6	559	5	Q01506	Q01506 plasmodium
45	101.5	6.6	559	5	Q01508	Q01508 plasmodium

## ALIGNMENTS

### RESULT 1

Q96P02 ID Q96P02 PRELIMINARY: PRT; 368 AA.  
 AC Q96P02;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE ANTHRAX TOXIN RECEPTOR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bradley K.A., Mogridge J., Mourez M., Collier R.J., Young J.A.T.;  
 RT "Identification Of the Cellular Receptor for Anthrax Toxin."  
 RL Nature 414:0-0(2001).  
 DR EMBL; AF421380; AAL26496.1; --  
 KW Receptor.  
 SQ SEQUENCE 368 AA; 41157 MW; 8A87B13FFA7D8753 CRC64;

Query Match	100.0%	Score 1536;	DB 4;	Length 368;
Best Local Similarity	100.0%	Pred. NO. 4.6e-127;		
Matches 295;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy 1	GOGGRRDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIYFS	60		
Db 27	GOGGRRDGGPACYGFDLYFILDKSGSVLHHWNEIYFVEQLAHKFISPOLRMSFIYFS	86		
Qy 61	TRGTTMLKLTDEQIROGLELQVLPGGDTYMHGFEERASEQIYYENRQGYRTASVII	120		
Db 87	TRGTTMLKLTDEQIROGLELQVLPGGDTYMHGFEERASEQIYYENRQGYRTASVII	146		
Qy 121	ALTDGELHEDLFYFSERANRSRDIGAIVCVGVKDENETQLARIADSKDHVPVNDGFG	180		
Db 147	ALTDGELHEDLFYFSERANRSRDIGAIVCVGVKDENETQLARIADSKDHVPVNDGFG	206		
Qy 181	ALQGIHSILKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVDRLVLCFSKINDSVT	240		

Query Match	37.4%	Score	574	DB	4	Length	245
Best Local Similarity	49.2%	Prd.	No. 1.4e-42				
Matches	120	Conservative	45	Mismatches	67	Indels	12
Gaps							
QY	52	LRMSFIVSTGRGTTLMKLTDEQIRQGLEBELQKVLPGGDTYMHGFERASEQIYYENRQ	111				
DB	1	MLRSFVSSQATIIPLTGDRKISKGLEDKRVPVGTYIHGGLKLANEQI--QKAG	58				
QY	112	GYRTASVITALTDGELHEDLFFYSEREANRSRDLCGAVYCVGKDFNETQIARIADSKDH	171				
DB	59	GLKTSIIIALTDGDKLUGLVPYSYAKKEAKISRLSGASYCVGVLDPEQALERIADSKQ	118				
QY	172	VFPVNDGFQALQGIHHSILKSKICIELAAEPSTCAGESFOVYVRGNGFRRHARNVDRVLC	231				
DB	119	VFPVKGSGFQALKGIIINSILAQSCTEILELPQSSVCGVEEQIVLSGRGFMGLSRNSVLC	178				
QY	232	SFKINDSVTLNEKPFPSVEDTYLLCAPILKEVGMKAALQVSMNDGLFSISSV-IITTH	290				
DB	179	TYTVNTEYTTTSVAPSVQLNSMLCPAPILNKAG-----ENGLTVTQAGVKWHLTH	229				
QY	291	CSDG	294				
DB	230	CTFG	233				
RESULT	3						
Q96EC6		PRELIMINARY					
ID	Q96EC6		PRT				97 AA.
AC	Q96EC6						
DT	01-DEC-2001	(TrEMBLrel. 19, Created)					
DT	01-DEC-2001	(TrEMBLrel. 19, Last sequence update)					
DT	01-DEC-2001	(TrEMBLrel. 19, Last annotation update)					
DE	UNKNOWN	(PROTEIN FOR IMAGE:4705862) (FRAGMENT).					
OS	Homo sapiens (Human)						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.						
NCBI	NCBI_TaxID:9606						

	Query Match	10.4%	Score 159.5;	DB 11;	Length 1161;	
	Best Local Similarity	27.9%;	Pred. No. 3.6e-05;			
	Matches	63;	Conservative 41;	Mismatches 93;	Indels 29;	Gaps 11;
QY	18	DLYFILDKSGSV-LHHWNIIYVFQLAHKFISPOLRMFSFVIFSTRGNTLMKLTE-----	71			
		:   :       :	:	:	:	:
DB	152	DIAFLDGGSGINQRDAQMDFVKALMGCEFASTLFLSMOYSNILKTHFTTFEKNIL	211			
		:   :       :	:	:	:	:
QY	72	DRQIRIQGLEELQKVLPGGDITYMHGEFRASQIYYENRQGVRTA-SVIALTDCGLHED	130			
		:   :       :	:	:	:	:
DB	212	DPOSLLVDPIVLQ-----GLTYTATGIRTVMEELFHSKNGSRKSARKILLVITDQCKNR	266			
		:   :       :	:	:	:	:
QY	131	LFFYSE--REANRSRLGAIVCVGVKD-FNE-TOLARI-----ADSKDHVPVNDGFQA	181			
		:   :       :	:	:	:	:

Query Match	10.4%	Score 159.5;	DB 11;	Length 1161;
Best Local Similarity	27.9%;	Pred. No. 3.6e-05;		
Matches	63;	Conservative 41;	Mismatches 93;	Indels 29; Gaps 11;
QY	18	DLYFLDKGSV-LHHWNIYVFEQLAHKFISPOLRMSFVIFSTRGVTLMKLTE-----	71	
		: :       : : :     :   :   :   :		
Db	152	DIAFLDGSINQRDFAQMKDFVKALMGFEFASTFLSLMOYSNILKTHFTTFEKNIL	211	
		: :       : : :     :   :   :   :		
QY	72	DREQIROGLEELQVLPGGDTVMHGEFERSBQIYYENRQGYRTA-SVLIATDGEELHD	130	
		: : :   :     : : :   : : :   : :   :   :		
Db	212	DPOSILVDPIVLQ-----GLTYTATGIRTVMEEELFHSKNGSRSAKKILLVITDGOKYRD	266	
		: : :   :     : : :   : : :   : :   :   :		
QY	131	LFIFYSE--REANKSRDLGAIYCVGVKD-FNE-TOLARI-----ADSKDHVPVMDGFOA	181	
		: : :   : : :   :       :   :   :   :		

```
Db 267 PLEYSVDVIPAADKA--GIIRYAIGVDAFQEPETALKEINTIGSAPPQDHVKVGN-FAA 322
QY 182 LOGTHILKSCIEILAAEPSTICAGESQVYVYRGNGRHRNVQ 227
Db 323 LRSIQRLQEK-----IFAIEGTGRSSSFQHEMSQEGFSALTSD 364

RESULT 5
Q9BPQ8 PRELIMINARY; PRT; 1332 AA.
AC Q9BPQ8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INTEGRIN ALPHA HRI PRECURSOR.
GN HRI2GAL
OS Eukaryotia roretzi (Sea squirt).
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
OC Stolidobranchia; Pyuridae; Halocynthia.
OX NCBI_TaxID=7729;
RN [1]
RC SEQUENCE FROM N.A.
RC TISSUE=HEMOCYTE;
RX MEDLINE=21103187; PubMed=11160215;
RA Miyazawa S., Azumi K., Nonaka M.;
RT "Cloning and characterization of integrin alpha subunits from the
RT solitary ascidian, Halocynthia roretzi.";
RL J. Immunol. 166:1710-1715(2001).
DR EMBL; AB048261; BAB21479.1; -.
DR HSP; P11215; IABX.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF01839; FG-GAP; 5.
DR Pfam; PF00357; Integrin_A; 1.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00327; VWA; 1.
DR PROSITE; PS50234; VWFA; 1.
KW Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1332 INTEGRIN ALPHA HRI.
SQ SEQUENCE 1332 AA; 145852 MW; 0D9108D2B05CFEAE CRC64;

Query Match 10.08; Score 153.5; DB 5; Length 1332;
Best Local Similarity 24.1%; Pred. No. 0.00015;
Matches 63; Conservative 42; Mismatches 81; Indels 75; Gaps 10;

QY 3 GGRREDGG----PACXGGFDLYFLDKSGSVLHHWNEIYFVEQLAHKFISPOLRWKSF 57
Db 188 GNRKESRNTECP--SGVDVLFVLOGSGVGRNFKVDWVRNIT----- 232
QY 58 VFSTRGTMLKLTEDRQIROGLEELQKVLPG-----GDTYMHGCFERAS 102
Db 233 -----AKLDIGKEIVRVGVVQYSHYVEGKINKQYITTEISIGEFKLLDNFENAV 283
QY 103 EQIYYENRQGYRT-----ASVIALTDGELHEDLFPYSREANRS 142
Db 284 DRI---QLQGYTTTGRALQKVIKIRDFDDAYIGNKQVLLLLTDGQAKONKILP--NANRL 338
QY 143 RDLGAIYVCVKDFNETOLARIA--DSKDHVPVNDGFOALOGIHSILKSCIETLA 199
Db 339 RNKGIAFVAGVGVDISELKLASGTDSDRVTFTVD-FGELDSIVKSLQTEIQSFVLE 397
QY 200 AEPSTICAGESQVYVYRGNGR 220
Db 398 GGKSAKTAG--YEMHFGENG 416

RESULT 6
Q04588
```

```
ID Q04588 PRELIMINARY; PRT; 724 AA.
AC Q04588;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MAJOR ANTIGEN HOMOLOGOUS SEQUENCE (EMP100).
OS Eimeria maxima.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
OX Eimeria.
OX NCBI_TaxID=5804;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=93149203; PubMed=8426611;
RA Pasamontes L.E., Hug D., Huembelin M., Weber G.;
RT "Sequence of a major Eimeria maxima antigen homologous to the Eimeria
RT tenella microneme protein Etp100.";
RL Mol. Biochem. Parasitol. 57:171-174(1993).
DR EMBL; M99058; AAA29076.1; -.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00090; tsp_1; 6.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00209; TSPI; 6.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50092; TSPI; 5.
DR PROSITE; PS50234; VWFA; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 724 AA; 75808 MW; AC2A0E7A346A7B9E CRC64;

Query Match 9.08; Score 139; DB 5; Length 724;
Best Local Similarity 26.6%; Pred. No. 0.0012;
Matches 59; Conservative 40; Mismatches 89; Indels 34; Gaps 14;

QY 13 CYGGFDLYFLDKSGSV-LHHWNEIYFVEQLAHKF-ISP-QLRMSFIVSTRGTTMLKL 69
Db 42 CTRLLDVLVYVDESGISCTSYGKVSFISNFACTMPLSPDDVRVGLVTFGTSATVRMDL 101
QY 70 TEDREIQROGLEELQKVLPG--GGDTYMHGCFERASEQIYYENRQGYR--TASVIALTDG 125
Db 102 SDSRAQNADLLAAAKKLPYANGSTYTHLGLAKA-EELFSFGGGRDNAPKMLVWTDG 160
QY 126 ELHEDLFFYSERE-----ANRSQDLGAIVYCVGV-KDFNETOLARIA--DSKDHV-PP- 174
Db 161 A-----SSRRQTLSSAAEKLNRGVIIIVLGVTGVSNAECRSIAGCSDTSDVECPR 212
QY 175 -VNDGFOALOGIHSILKSCIEI-----LAAEPSTI--CAGE 209
Db 213 YLQSNWGGVSSQINGIIKAACKDLAKDAVCSEWSEYGPCEGE 254

RESULT 7
Q043981 PRELIMINARY; PRT; 712 AA.
AC Q043981;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MICRONEME PROTEIN ETMIC-1 PRECURSOR.
OS MIC-1.
OS Eimeria tenella.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
OX Eimeria.
OX NCBI_TaxID=5802;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=HOUGHTON;
RX MEDLINE=92131064; PubMed=1775171;
RA Tomley F.M., Clarke L.E., Kawazoe U., Dijkema R., Kok J.J.;
RT "Sequence of the gene encoding an immunodominant microneme protein of
RT Eimeria tenella.";
RL Mol. Biochem. Parasitol. 49:277-288(1991).
RN [2]
```

[illegible]

RA	Strausberg R.:			
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE			
CC	TRYPSIN FAMILY.			
DR	EMBL;	AF109906;	AAC84162.1;	-
DR	EMBL;	AF049850;	AAC05284.1;	-
DR	EMBL;	BC011086;	AAH11086.1;	-
DR	HSSP;	P00763;	LEPT.	
DR	MGD;	MGI:88226;	C2.	
DR	InterPro;	IPR001314;	Chymotrypsin.	
DR	InterPro;	IPR000436;	Sushi_SCR_CCP.	
DR	InterPro;	IPR001254;	Trypsin.	
DR	InterPro;	IPR002035;	WVFA.	
DR	Pfam;	PF00084;	sushi; 2.	
DR	Pfam;	PF00089;	trypsin; 2.	
DR	Pfam;	PF00092;	vwa; 1.	
DR	PRINTS;	PR00722;	CHYMOTRYPSIN.	
DR	PRINTS;	PR00453;	VWFADOMAIN.	
DR	SMART;	SM00032;	CCP; 2.	
DR	SMART;	SM00020;	Tryp-Spc; 1.	
DR	SMART;	SM00327;	VWA; 1.	
DR	PROSITE;	PS50240;	TRYPSIN_DOM; 2.	
DR	PROSITE;	PS00134;	TRYPSIN_HIS; UNKNOWN_1.	
DR	PROSITE;	PS00135;	TRYPSIN_SER; UNKNOWN_1.	
DR	PROSITE;	PS50234;	WVFA; 2.	
KW	Hydrolase; Serine protease.			
SEQUENCE	760 AA;	84741 MW;	91C896A3EDC7D448	CRC64;
Query Match	8.7%;	Score 134;	DB 11;	Length 760;
Best Local Similarity	23.5%;	Pred. No. 0.0035;		
Matches	75;	Conservative	55;	Mismatches 115; Indels 74; Gaps 16;
QY	15	GGFDLAFILDKSGVSLVHNEIY-YFVEOLAHKFISPOLRMS--FIVFSTRGTTMLKLTE	71	
Db	258	GHUNLYLLLDASQSVTEKDFDKFKSAELMVERIFSEFVNSVAITTFASQPKTMSILS	317	
QY	72	DREQIRQGLEELQKVLPGGTYMHGFEFRASEQIYE-----NRQGYRTAS-	117	
Db	318	ERSQ-----DVTEFISLDSASYKDHENATGNTVEILRVYSMMQSDRLGMETSAW	371	
QY	118	-----VTIALTDELHEDLFFYSEREANRSRDIGAI-----YVCVGV-----KD	156	
Db	372	KEIRHTILLTDGK--SNMGDSPKAVTRIRELLSIEQNRDDVDLYIAGVGKLDYDWKE	429	
QY	157	FNETQLARIADSKDHVPVNDGFOALOGIIHSLKKSCIEILLAAEPTICAGESFOVVVR	216	
Db	430	LNE--LGSKDKGERHAFILODA-KALQIQEIHMLDVSKL-----TDTICG-----V	472	
QY	217	NGGFHARNVDRV--LCSEKINDSVTLNEKPFVSVEDTYLLCPAPILKEGVMK--AALQVS	272	
Db	473	GNMSANASQDERTPWQVTFPKSKETCGS--LISDQVLTAAHCFHDHIDHHLWRVN	530	
QY	273	MND-----GLSFISSVII	286	
Db	531	VGDPTSQHGKEELVEDVII	549	
RESULT	9			
Q9U8J9				
ID	Q9U8J9	PRELIMINARY;	PRT;	765 AA.
AC	Q9U8J9;			
DT	01-MAY-2000	(TrEMBLrel. 13, Created)		
DT	01-MAY-2000	(TrEMBLrel. 13, Last sequence update)		
DT	01-DEC-2001	(TrEMBLrel. 19, Last annotation update)		
DE	THROMBOSPONDIN-RELATED ADHESIVE PROTEIN HOMOLOG.			
GN	MIC2.			
OS	Neospora caninum.			
OC	Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae;			
OC	Neospora.			
OX	NCBI_TaxID=29176;			
RN	[1]			
RP	SEQUENCE FROM N.A.			

FT	NON_TER	517	517
SQ	SEQUENCE	517 AA; 56727 MW; 9B6972F44A1BD88F	CRC64;
Query Match 8.6%; Score 131.5; DB 4; Length 517;			
Best Local Similarity 29.6%; Pred. No. 0.0035;			
Matches 61; Conservative 39; Mismatches 77; Indels 29; Gaps			
QY	13	CYGGFDL---YFILDKSGSV-LHHWNEIYYFYVQLAHKF-ISP-OLRMSFIVEST---	RG 63
Db	133	CSRGVDIKADMFLVDGYSIGIANFKVKRAFLVLVKSFESIPNRVQISLVQYSRDPHTE	192
QY	64	TTLMKLTEDREQIROGLEELQKVLPGDGYMHGEFERASEIOIYYENROGYRT--ASVI	119
Db	193	FLLKKFTKVEDI-----EAINTPPYRGSTNTGKAMYVREKIFVPSP-GSRSNPKVM	246
QY	120	IALTGDGLHDLFFYSERANRSRDGAIVYCVGVKDFNETOLARIAD--SKDHVFPVND	177
Db	247	ILITDGK-SSDAF---RDPAILKRNDSVEIFAVGVKDAVDSELEAIASPETHVTVED	302
QY	178	GFOALQGIIHSILKKSCIEI---LAA 200	
Db	303	-FOAFQRISFELTQSLCIRIEQLAA 327	
RESULT 11			
ID	Q9UGC3	PRELIMINARY;	PRT; 660 AA.
AC	Q9UGC3;		
DT	01-MAY-2000	(TtEMBLrel. 13, Created)	
DT	01-MAY-2000	(TtEMBLrel. 13, Last sequence update)	
DT	01-DEC-2001	(TtEMBLrel. 19, Last annotation update)	
DE	DJ234P15.1	(COLLAGEN, TYPE XII, ALPHA 1) (FRAGMENT).	
GN	COL12A1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OX	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
NCBI_TaxID	9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Smith M.;		
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AL080250; CAB65984.1; "		
DR	HSP; P11215; 1BHQ.		
DR	InterPro; IPR003961; FN_III.		
DR	InterPro; IPR002035; vWFA.		
DR	Pfam; PF00041; fn3; 2.		
DR	Pfam; PF00092; vwa; 2.		
DR	PRINTS; PR00453; VWFADOMAIN.		
DR	SMART; SM00060; FN3; 2.		
DR	SMART; SM00327; vWA; 2.		
DR	PROSITE; PS50234; vWFA; 2.		
KW	Collagen.		
FT	NON_TER	660	660
SQ	SEQUENCE	660 AA; 72523 MW; 3E3A10A285ECAA51	CRC64;
Query Match 8.5%; Score 131; DB 4; Length 660;			
Best Local Similarity 29.8%; Pred. No. 0.0054;			
Matches 59; Conservative 38; Mismatches 75; Indels 26; Gaps			
QY	18	DLXFLDKSGSV-LHHWNEIYYFYVQLAHKF-ISP-OLRMSFIVEST--RGTTLMKLTE	71
Db	469	DIVFLVDGYSIGIANFKVKRAFLVLVKSFESIPNRVQISLVQYSRDPHTETLKAKTK	528
QY	72	DREQIROGLEELQKVLPGDGYMHGEFERASEIOIYYENROGYRT--ASVIALTDLGEL	127
Db	529	VEDII-----EAINTPPYRGSTNTGKAMYVREKIFVPSP-GSRSNPKVMILLTDGK-	581
QY	128	HEDLFFYSERANRSRDGAIVYCVGVKDFNETOLARIAD--SKDHVFPVNDGFOALQGI	185
Db	582	SSDAF---RDPAILKRNDSVEIFAVGVKDAVRSELEAIASPPAEETHVTVED-FDAFOR	637

RC TISSUE-MAMMARY TUMOR. WAP-TGF ALPHA MODEL. 7 MONTHS OLD, CROSS  
RC TISSUE.;  
RA Strausberg R.;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC003908; AAH03908.1; -;  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 221 AA; 24505 MW; 31AB702B42E9CEB3 CRC64;

Query Match 8.3%; Score 128; DB 11; Length 221;  
Best Local Similarity 44.2%; Pred. No. 0.0022;  
Matches 23; Conservative 12; Mismatches 17; Indels 0; Gaps 0;

OY 243 EXPFSVEDTYLLCPAPILKEVGMKAALQVMNDGLSFSSSVITTHCHSGD 294  
||| |:::|||||:| : |||||:| |::|  
Db 1 EXPVSIQPSIILCAPVLNKGDTLEIVSYNDGKSAVSRLTITATECTNG 52

RESULT 14  
O9UAL13 PRELIMINARY; PRT: 3767 AA.

ID O9UAL13 Q21340; AC  
AC O9UAL13; Q21340; (Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE TRANSMEMBRANE CELL ADHESION RECEPTOR MUA-3 PRECURSOR (K08E5.3 PROTEIN)  
DE (FRAGMENT)  
GN MUA-3 OR K08E5.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditoidea;  
OC Rhabditiidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N2;  
RA Lu Z., Vogel B., Hedgecock E.;  
RT "mua-3 mRNA Splicing Pattern Revealed";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Kershaw J.K.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
investigating biology";  
RL Science 282:2012-2018(1998).  
DR EMBL; AF139060; AAD29428.1; -;  
DR EMBL; Z30974; CAAB3226.2; -;  
DR EMBL; Z30423; CAAB3226.2; JOINED.  
DR HSP; P01130; ILDL.  
DR InterPro; IPR000152; Asx\_hydroxyl.  
DR InterPro; IPR001762; Disintegrin.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001881; EGF-Ca.  
DR InterPro; IPR002172; LDL\_recept\_A.  
DR InterPro; IPR000082; SEA.  
DR InterPro; IPR002035; VWFA.  
DR Pfam; PF00008; EGF; 33.  
DR Pfam; PF00057; ldl\_recept\_a; 4.  
DR Pfam; PF01390; SEA; 2.  
DR Pfam; PF00092; vwa; 1.  
DR PRINTS; PR00289; DISINTEGRIN.  
DR PRINTS; PR00261; LDLRECEPTOR.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00181; EGF; 51.  
DR SMART; SM00179; EGF\_CA; 1.  
DR SMART; SM00001; EGF\_like; 45.  
DR SMART; SM00192; LDLa; 4.  
DR SMART; SM00200; SEA; 2.

OY 18 DLYFILDKSGVLHHNEIYFYEQLAHKFI SP--QLRMSFIVESTRGTTLMLKTED-- 72  
|:|:| |::| |::| |::| |::| |:  
Db 17 DIAFLIDGSGI-----NR LDF---QRMKFEVSTVMGQFKSKTLFA-----LMQISEDYF 64

OY 73 -----REQIROGLEELQKLVLPGGDTYMHGEFERASEQIYYENRGYR- 114  
||| |::| |::| |::| |:  
Db 65 THFTFNDFKNPSPKLLVRPIRQLL-----GRHTATGIRKVVRLELFH-SKSGARE 114

OY 115 -TASVIIALTGELHEDLFFYSE--REANRSRDIGAIYCVGVKD-FN-----ETQLARIA 166  
::: ||||| |::| |::| |::| |:  
Db 115 NALKILVVITDGKFGDPGLGYEDVIPADRK--GVRIYVGVDGAFNSWKSREELNTIA 171

OY 167 DSK--DHVPFPYDFOALOGIIHSILKKSCIEILAAPPSTICAGESQQVYVRNGFRHA 223  
||||| |:| |::| |::| |::| |:  
Db 172 SKPCGDHVQVNN-PEAVKTIQNLOQKT----FAIECTQTGSTTSFECEMSQEGFSAA 225

RESULT 13  
O99L17 PRELIMINARY; PRT: 221 AA.

ID O99L17  
AC O99L17; Q21340; (Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 24.5 KDA PROTEIN (FRAGMENT).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.

Query Match 8.5%; Score 130.5; DB 6; Length 920;  
Best Local Similarity 26.4%; Pred. No. 0.0093;  
Matches 63; Conservative 41; Mismatches 72; Indels 63; Gaps 15;

OY 18 DLYFILDKSGVLHHNEIYFYEQLAHKFI SP--QLRMSFIVESTRGTTLMLKTED-- 72  
|:|:| |::| |::| |::| |::| |:  
Db 17 DIAFLIDGSGI-----NR LDF---QRMKFEVSTVMGQFKSKTLFA-----LMQISEDYF 64

OY 73 -----REQIROGLEELQKLVLPGGDTYMHGEFERASEQIYYENRGYR- 114  
||| |::| |::| |::| |:  
Db 65 THFTFNDFKNPSPKLLVRPIRQLL-----GRHTATGIRKVVRLELFH-SKSGARE 114

OY 115 -TASVIIALTGELHEDLFFYSE--REANRSRDIGAIYCVGVKD-FN-----ETQLARIA 166  
::: ||||| |::| |::| |::| |:  
Db 115 NALKILVVITDGKFGDPGLGYEDVIPADRK--GVRIYVGVDGAFNSWKSREELNTIA 171

OY 167 DSK--DHVPFPYDFOALOGIIHSILKKSCIEILAAPPSTICAGESQQVYVRNGFRHA 223  
||||| |:| |::| |::| |::| |:  
Db 172 SKPCGDHVQVNN-PEAVKTIQNLOQKT----FAIECTQTGSTTSFECEMSQEGFSAA 225

RESULT 13  
O99L17 PRELIMINARY; PRT: 221 AA.

ID O99L17  
AC O99L17; Q21340; (Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 24.5 KDA PROTEIN (FRAGMENT).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.

Query Match 8.5%; Score 130.5; DB 6; Length 920;  
Best Local Similarity 26.4%; Pred. No. 0.0093;  
Matches 63; Conservative 41; Mismatches 72; Indels 63; Gaps 15;

OY 18 DLYFILDKSGVLHHNEIYFYEQLAHKFI SP--QLRMSFIVESTRGTTLMLKTED-- 72  
|:~::~| |::| |::| |::| |:  
Db 17 DIAFLIDGSGI-----NR LDF---QRMKFEVSTVMGQFKSKTLFA-----LMQISEDYF 64

OY 73 -----REQIROGLEELQKLVLPGGDTYMHGEFERASEQIYYENRGYR- 114  
||| |::| |::| |::| |:  
Db 65 THFTFNDFKNPSPKLLVRPIRQLL-----GRHTATGIRKVVRLELFH-SKSGARE 114

OY 115 -TASVIIALTGELHEDLFFYSE--REANRSRDIGAIYCVGVKD-FN-----ETQLARIA 166  
::: ||||| |::| |::| |::| |:  
Db 115 NALKILVVITDGKFGDPGLGYEDVIPADRK--GVRIYVGVDGAFNSWKSREELNTIA 171

OY 167 DSK--DHVPFPYDFOALOGIIHSILKKSCIEILAAPPSTICAGESQQVYVRNGFRHA 223  
||||| |:~::~| |::| |::| |::| |:  
Db 172 SKPCGDHVQVNN-PEAVKTIQNLOQKT----FAIECTQTGSTTSFECEMSQEGFSAA 225

RESULT 13  
O99L17 PRELIMINARY; PRT: 221 AA.

ID O99L17  
AC O99L17; Q21340; (Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 24.5 KDA PROTEIN (FRAGMENT).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.

Query Match 8.5%; Score 130.5; DB 6; Length 920;  
Best Local Similarity 26.4%; Pred. No. 0.0093;  
Matches 63; Conservative 41; Mismatches 72; Indels 63; Gaps 15;

OY 18 DLYFILDKSGVLHHNEIYFYEQLAHKFI SP--QLRMSFIVESTRGTTLMLKTED-- 72  
|:~::~| |::| |::| |::| |:  
Db 17 DIAFLIDGSGI-----NR LDF---QRMKFEVSTVMGQFKSKTLFA-----LMQISEDYF 64

OY 73 -----REQIROGLEELQKLVLPGGDTYMHGEFERASEQIYYENRGYR- 114  
||| |::| |::| |::| |:  
Db 65 THFTFNDFKNPSPKLLVRPIRQLL-----GRHTATGIRKVVRLELFH-SKSGARE 114

OY 115 -TASVIIALTGELHEDLFFYSE--REANRSRDIGAIYCVGVKD-FN-----ETQLARIA 166  
::: ||||| |::| |::| |::| |:  
Db 115 NALKILVVITDGKFGDPGLGYEDVIPADRK--GVRIYVGVDGAFNSWKSREELNTIA 171

OY 167 DSK--DHVPFPYDFOALOGIIHSILKKSCIEILAAPPSTICAGESQQVYVRNGFRHA 223  
||||| |:~::~| |::| |::| |::| |:  
Db 172 SKPCGDHVQVNN-PEAVKTIQNLOQKT----FAIECTQTGSTTSFECEMSQEGFSAA 225

RESULT 13  
O99L17 PRELIMINARY; PRT:

```
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_32.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 6.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS00068; LDLRA_2; 2.
DR PROSITE; PS00068; LDLRA_2; 2.
DR PROSITE; PS00024; SEA; 4.
DR PROSITE; PS00234; VWFA; 2.
KW Calcium-binding; EGF-like domain; Glycoprotein; Receptor; Repeat;
KW SIGNAL.
FT SIGNAL. 1 24 POTENTIAL.
FT CHAIN 25 >3767 TRANSMEMBRANE CELL ADHESION RECEPTOR MUA-
FT 3.
FT NON_TER 3767 3767
FT SEQUENCE 3767 AA; 417284 MW; 8DA3AE5EA50AEB8E CRC64;

Query Match 8.3%; Score 128; DB 5; Length 3767;
Best Local Similarity 27.1%; Pred. No. 0.11;
Matches 64; Conservative 31; Mismatches 75; Indels 66; Gaps 13;

Qy 13 CYGCF-----DLYFILDKSGSVLHH--WNEIYFVEQLA 44
Dy 1199 CYGFEVDVSSNANLPPGRVCTVQTCPKQTDLVFLDGGSGISGYVFKNEVLFVREFV 1258
Qy 45 HKFI--SPQLRMSFIVFSTGTTLMLKTE--DREIQRGLEELQKVLPG-----GDTYMH- 95
Dy 1259 ELFEIGRSKRVGLIQYSDQIRHEFDLDQYGRDLSLKGISETO-YLTGLTRTGAATQHM 1317
Qy 96 --EGFERASEQIYYENRQGYR-----TASVITALTGDELHEDLFYSEREANRSRDILGAI 148
Dy 1318 VOEGF-----SERRGAPQOSDIARVAIILTDGRSQDNV----TGPADSARKLSIN 1364
Qy 149 YVCVGKVD-FNETQLARIADSKDHVPVNDGFGQALQGIHHSILKSKCIEILAEPS 203
Dy 1365 TFAIGVTDHVLASELSIAGSPNRWFYV-DKFKDLTLRLSRMIQK-----AACPS 1413

RESULT 15
Q9J130 PRELIMINARY; PRT; 1151 AA.
AC Q9J130;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INTEGRIN BETA 2 ALPHA SUBUNIT.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Fathallah D.W. Sr., Zerria K. Jr.;
RT "Cloning of the rat cd11b cDNA sequence.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF268593; AAF81280.1;
DR HSSP; P11215; 1A8X.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF01839; FG-GAP; 5.
DR Pfam; PF00357; Integrin_A; 1.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00327; Int_alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS0234; VWFA; 1.
SQ SEQUENCE 1151 AA; 126943 MW; 8F7856595D4074CA5 CRC64;
```

Query Match 8.0%; Score 123; DB 11; Length 1151;  
Best Local Similarity 23.9%; Pred. No. 0.058;

```
Matches 70; Conservative 46; Mismatches 105; Indels 72; Gaps 15;

Qy 18 DLYFILDKSGSVLHHWNEIYFVEQLAHKETS---PQLRMSFIVFSTGTTLMLKLTED-- 72
Dy 150 NIAFLIDGSGSI---NTIDF---QKMKFVSTVMDQFQSKTLFS-----LMQYSDEF 197
Qy 73 ---REQIRQGLEELQKVLPG---GGDTYMHGEPERASEQIYYE-NRQGYRTASVITALT 123
Dy 198 THFTFNDFKRNPDPPKSHVPIRLQNGRTKTASGIRKVVVRELFOKINGCARDNAAKILVVIT 257
Qy 124 DGEIHEDLFYSEREANRSRDILGAIYVCVGKVD-FNETQLARIAD-----SKDHVFPVN 176
Dy 258 DGEKFGDPLNY-EDVIPEAEAGIIRYVIGVGNAFHFKPQSRRELDTIASKPAGDHVFQV- 315
Qy 177 DGFQALQGIHHSILKSKCIEILAEPSITICAGESFOVVVRGNGFRHARNVDRLVLCSPKIN 236
Dy 316 DNFPALNTIRNQLOEK-----IFAIEGTOTGSTSFHEHMSOEGFSA----- 358
Qy 237 DSVTLNKPFESVEDTYLLCPAPILKEVG---MKAALQVSNMNDGLSFISSVI 285
Dy 359 --ITSN-----CPLLGSVSGSDWAGGAFLYPSKDKKASFINTRI 395
```

Search completed: August 9, 2002, 10:46:21  
Job time: 844 sec

us-09-970-076-2\_copy\_27\_321.rspt

Fri Aug 9 10:57:03 2002

\_\_\_\_\_



GenCore version 4.3  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 8, 2002, 22:10:27 ; Search time 109.68 seconds

(without alignments)

3166.719 Million cell updates/sec

Title: US-09-970-076-1

Perfect score: 1414

Sequence: 1 eggaccgcgaggaaggcc.....aaaaaaaaaaaaaaaaaaaa 1414

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/2/ina/PCrUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51.6	3.6	2158	1	US-07-602-608-1
2	51.6	3.6	2158	1	US-08-261-578-1
3	50.8	3.6	3829	2	US-08-631-097-8
4	50.8	3.6	3829	4	US-08-810-712-6
5	49.8	3.5	658	4	US-08-998-416-595
6	48.4	3.4	1898	1	US-08-342-411A-1
7	47	3.3	1279	3	US-09-248-335-25
8	47	3.3	7218	1	US-08-232-463-14
9	46.8	3.3	1364	1	US-08-265-087-3
10	46.8	3.3	1364	1	US-08-621-493-3
11	46.8	3.3	1364	2	US-08-965-688-3
12	46.8	3.3	1364	4	US-09-260-173-3
13	46.8	3.3	1813	5	PCR-US94-12863-3
14	46.8	3.3	2550	6	5258287-23
15	46.6	3.3	1776	3	US-08-655-352-10
16	46.6	3.3	1825	4	US-09-461-697-75
17	46.4	3.3	1098	3	US-09-248-335-35
18	46.4	3.3	3214	1	US-08-484-105-17
19	46.4	3.3	3214	1	US-08-484-106-17
20	46	3.3	760	1	US-08-276-452A-49
21	46	3.3	760	2	US-08-798-744-49
22	46	3.3	1512	2	US-08-909-965C-8
23	46	3.3	1600	4	US-07-861-458C-37
24	46	3.3	1607	6	5196333-3
25	46	3.3	2096	3	US-09-008-481A-10
26	46	3.3	2096	3	US-09-195-656A-16
27	46	3.3	2096	4	US-09-309-592-10

28	45.8	3.2	1023	1	US-08-252-966B-16	Sequence 16, Appl
29	45.6	3.2	1454	4	US-09-372-422A-19	Sequence 19, Appl
30	45.6	3.2	1604	1	US-08-665-966-9	Sequence 9, Appl
31	45.6	3.2	1604	3	US-09-041-780-9	Sequence 9, Appl
c 32	45.6	3.2	3138	1	US-07-867-106-4	Sequence 4, Appl
33	45.4	3.2	1736	3	US-09-182-816-22	Sequence 22, Appl
c 34	45.4	3.2	1736	3	US-09-182-816-24	Sequence 24, Appl
35	45.4	3.2	1736	3	US-09-471-528-22	Sequence 22, Appl
c 36	45.4	3.2	1736	3	US-09-471-528-24	Sequence 24, Appl
37	45.4	3.2	1736	4	US-09-634-530-22	Sequence 22, Appl
c 38	45.4	3.2	1736	4	US-09-634-530-24	Sequence 24, Appl
39	45.2	3.2	966	1	US-08-514-014-7	Sequence 7, Appl
40	45.2	3.2	966	2	US-08-833-823-7	Sequence 7, Appl
41	45.2	3.2	1386	2	US-08-897-340-1	Sequence 1, Appl
42	45.2	3.2	1386	3	US-09-252-329-1	Sequence 1, Appl
43	45.2	3.2	1540	3	US-08-977-001-2	Sequence 2, Appl
44	45.2	3.2	2026	2	US-08-993-228-3	Sequence 3, Appl
45	45	3.2	882	2	US-08-909-965C-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1  
US-07-602-608-1  
; Sequence 1, Application US/07602608  
; Patent No. 5382524  
; GENERAL INFORMATION:  
; APPLICANT: Desnick, Robert J.  
; APPLICANT: Bishop, David F.  
; APPLICANT: Ioannou, Yiannis A.  
; APPLICANT: Wang, Anne M.  
; TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY  
; TITLE OF INVENTION: ACTIVE ALPHA-N-ACETYL GALACTOSAMINIDASE  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/602,608  
; FILING DATE: 24-OCT-1990  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 6923-008  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2158 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 345..1580  
US-07-602-608-1

Query Match 3.6%; Score 51.6; DB 1; Length 2158;

Query Match 3.6%; Score 51.6; DB 1; Length 2158;  
Best Local Similarity 62.3%; Pred. No. 0.00038;-  
Matches 81; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

CLONE: not applicable  
LIBRARY: not applicable  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: not applicable  
MAP POSITION: not applicable

APPLICANT: Mohr, Christine  
 APPLICANT: Wendland, Jurgén  
 APPLICANT: Knechtle, Philipp  
 APPLICANT: Reibschung, Corinne  
 TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYYPIL  
 TITLE OF INVENTION: AND USES THEREOF  
 NUMBER OF SEQUENCES: 1152  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: No. 6239264artis Corporation  
 STREET: 3054 Cornwallis Road  
 CITY: Research Triangle Park  
 STATE: No. 6239264th Carolina  
 COUNTRY: USA  
 ZIP: 27709

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/998,416  
 FILING DATE: 24-DEC-1997  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: CH 0016/97  
 FILING DATE: 31-DEC-1996  
 ATTORNEY/AGENT INFORMATION:

NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38, 241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 595:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 658 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1408RP  
8-998-416-595

[illegible]

RESULT 6  
US-08-342-411A-1  
; Sequence 1, Application US/08342411A  
; Patent No. 5639616  
; GENERAL INFORMATION:  
; APPLICANT: LIAO, Shutsung  
; APPLICANT: SONG, Ching

us-09-970-076-1.rni

Fri Aug 9 10:56:52 2002

```

; TITLE OF INVENTION:  UBQUITOUS NUCLEAR RECEPTOR:
; TITLE OF INVENTION:  COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES:  36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  Arnold, White & Durkee
; STREET:  P.O. Box 4433
; CITY:  Houston
; STATE:  TX
; COUNTRY:  USA
; ZIP:  77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Floppy disk
; OPERATING SYSTEM:  PC-DOS/MS-DOS
; SOFTWARE:  PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:  US/08/342,411A
; FILING DATE:  18-NOV-1994
; CLASSIFICATION:  435
; ATTORNEY/AGENT INFORMATION:
; NAME:  KITCHELL, BARBARA S.
; REGISTRATION NUMBER:  33,928
; REFERENCE/DOCKET NUMBER:  ARCD154
; TELEPHONE:  (512) 418-3000
; TELEFAX:  (713) 789-2679
; TELEX:  79-0924
; INFORMATION FOR SEQ ID NO:  1:
; SEQUENCE CHARACTERISTICS:
; LENGTH:  1898 base pairs
; TYPE:  nucleic acid
; STRANDEDNESS:  single
; TOPOLOGY:  linear
; FEATURE:
; NAME/KEY:  CDS
; LOCATION:  71..1450
; US-08-342-411A-1

```

```

Query Match      3.4%; Score 48.4; DB 1; Length 1898;
Best Local Similarity 68.4%; Pred. No. 0.0025;
Matches 67; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1316 ctttctctagttccgtattcgaatcccgagttgtctaacattcaataagtagctatat 1375
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1793 CTCCTCTCTGTTTATTATATAAACTTAAACAGAAAAAATATATATATATATATATAT 1852
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1376 gaaatcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1413
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1853 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1890

```

```

RESULT 7
US-09-248-335-25
; Sequence 25, Application US/09248335
; Patent No. 6096504
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; TITLE OF INVENTION:  PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
; FILE REFERENCE:  CL-1128-A
; CURRENT APPLICATION NUMBER:  05/09/248,335
; EARLIER FILING DATE:  1999-02-10
; EARLIER APPLICATION NUMBER:  08/924,759
; EARLIER FILING DATE:  1997-September-05
; NUMBER OF SEQ ID NOS:  74
; SOFTWARE:  Microsoft word Version 7.0A
; SEQ ID NO 25
; LENGTH:  1279
; TYPE:  DNA
; ORGANISM:  maize
; US-09-248-335-25

```

```

Query Match      3.3%; Score 47; DB 3; Length 1279;
Best Local Similarity 74.7%; Pred. No. 0.0049;
Matches 59; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1336 attcaaatcccgagtgcttaacattcaataaataagctatgaaatcaaaaaaaaaaaaaa 1395
||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1183 attcaattccagtgctgtgaaaaaataaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1242

QY 1396 aaaaaaaaaaaaaaaaaaaaaa 1414
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1243 aaaaaaaaaaaaaaaaaaaaaa 1261

RESULT 8
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT:  DORNER, F.
; APPLICANT:  SCHEIFLINGER, F.
; APPLICANT:  FALKNER, F. G.
; TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES:  52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  Foley & Lardner
; STREET:  1800 Diagonal Road, Suite 500
; CITY:  Alexandria
; STATE:  VA
; COUNTRY:  USA
; ZIP:  22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Floppy disk
; OPERATING SYSTEM:  PC-DOS/MS-DOS
; SOFTWARE:  PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  US/08/232,463
; FILING DATE:
; CLASSIFICATION:  435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER:  US/07/935,313
; FILING DATE:
; APPLICATION NUMBER:  EP 91 114 300.6
; FILING DATE:  26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME:  BENT, Stephen A.
; REGISTRATION NUMBER:  29,768
; REFERENCE/DOCKET NUMBER:  30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:  (703)836-9300
; TELEFAX:  (703)683-4109
; TELEX:  899149
; INFORMATION FOR SEQ ID NO:  14:
; SEQUENCE CHARACTERISTICS:
; LENGTH:  7218 base pairs
; TYPE:  nucleic acid
; STRANDEDNESS:  single
; TOPOLOGY:  linear
; IMMEDIATE SOURCE:
; CLONE:  ptzgpt-fls
; US-08-232-463-14

```

```

Query Match      3.3%; Score 47; DB 1; Length 7218;
Best Local Similarity 6.7%; Pred. No. 0.012;
Matches 26; Conservative 199; Mismatches 164; Indels 0; Gaps 0;

QY 792 ccataatgtcgagagagtcattcaagtgtcgtgagagaaacaggttcgcacatgcc 851
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1011 CCATACGCTCACAGAAATTAATTCGAGCTTGGCTGCAGTCCGAGGAGCTTGCATYYY 1070
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 852 gcaacgtggacaggggtctctctgcagcttcaagatcaatgactcgtcacatcaatgaga 911

```



us-09-970-076-1.rni

Fri Aug 9 10:56:52 2002

```

; Patent No. 6168923
; GENERAL INFORMATION: Scott, Phillip
; APPLICANT: Trinchieri, Giorgio
; TITLE OF INVENTION: Compositions and Methods for Use of
; IL-12 as an Adjuvant
;
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, PO Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/260,173
; FILING DATE: 01-Mar-1999
; CLASSIFICATION: <Unknown>
; 17-JUN-1994
; 18-APR-1994
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/965,688
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/265,087
; FILING DATE: 17-JUN-1994
; APPLICATION NUMBER: US 08/229,282
; FILING DATE: 18-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST51AUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9206
; TELEFAX: 215-540-5818
;
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1364 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 101..859
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-260-173-3

Query Match 3.3%; Score 46.8; DB 4; Length 1364;
Best Local Similarity 73.2%; Pred. No. 0.0057; Indels 0; Gaps 0;
Matches 60; Conservative 0; Mismatches 22;

QY 1333 tgtattcaatcccgagtgtctaacattcaataaataagctatgaaatcaaaaaaa 1392
DB 1268 TGTATTGAAAAATATTTTAAAGTGTCTTAAATAAAGTATTGAATTAATAAAAAA 1327
QY 1393 aaaaaaa 1414
DB 1328 AAAAAA 1349

RESULT 13
PCT-US94-12883-3
; Sequence 3, Application PC/TUS9412883
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: UBIQUITOUS NUCLEAR RECEPTOR: COMPOSITIONS AND
; METHODS
;
; Patent No. 5976539
; GENERAL INFORMATION: Scott, Phillip
; APPLICANT: Trinchieri, Giorgio
; TITLE OF INVENTION: Compositions and Methods for Use of
; IL-12 as an Adjuvant
;
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, PO Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/965,688
; FILING DATE:
; CLASSIFICATION:
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/621,493
; FILING DATE: 25-MAR-1996
; APPLICATION NUMBER: 08/265,087
; FILING DATE: 17-JUN-1994
; APPLICATION NUMBER: US 08/229,282
; FILING DATE: 18-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST51AUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9206
; TELEFAX: 215-540-5818
;
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1364 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 101..859
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-08-965-688-3

Query Match 3.3%; Score 46.8; DB 2; Length 1364;
Best Local Similarity 73.2%; Pred. No. 0.0057; Indels 0; Gaps 0;
Matches 60; Conservative 0; Mismatches 22;

QY 1333 tgtattcaatcccgagtgtctaacattcaataaataagctatgaaatcaaaaaaa 1392
DB 1268 TGTATTGAAAAATATTTTAAAGTGTCTTAAATAAAGTATTGAATTAATAAAAAA 1327
QY 1393 aaaaaaa 1414
DB 1328 AAAAAA 1349

RESULT 12
US-09-260-173-3
; Sequence 3, Application US/09260173
```



Fri Aug 9 10:56:52 2002

Db 1636 TCTCTAACTTTCATGACATGCATTATTTATTTATCTACTCTCTAAAGATCCTT 1695  
QY 1334 gtattcaaatcccgctgtctaacattcaataaataagctatatgaaatcaaaaaa 1393  
Db 1696 TAAATTAATTCGGAAGCCTTTATGCTAAATAAAAAAAAAAAAAAAAAA 1755  
QY 1394 aaaaaaaaaaaaaaaaaa 1414  
Db 1756 AAAAAAAAAAAAAAAAAA 1776

Search completed: August 9, 2002, 00:23:01  
Job time: 7954 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:47:01 ; Search time 31.3 Seconds  
(without alignments)  
697.694 Million cell updates/sec

Title: US-09-970-076-6  
Perfect score: 3025  
Sequence: 1 MATERRALGIGFQWLSLAT.....QAPPPNRPSPRPPRPSV 564

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5  
105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3025	100.0	564	1 ATR_HUMAN	Q9h6x2 homo sapien
2	2901	95.9	562	1 ATR_MOUSE	Q9cz52 mus musculus
3	1060	35.0	386	1 CMG2_HUMAN	P58335 homo sapien
4	200.5	6.6	3124	1 CA1C_CHICK	P13944 gallus gall
5	199	6.6	485	1 SSGP_VOLCA	P21997 volvox eart
6	197.5	6.5	141	1 YPRO_OWEFU	P21260 owenta fusi
7	191.5	6.3	306	1 EXTN_DAUCA	P06599 gaurus caro
8	188.5	6.2	1206	1 FMN1_MOUSE	Q05859 mus musculus
9	188.5	6.2	1468	1 FMN1_MOUSE	Q05860 mus musculus
10	187.5	6.2	555	1 GP1_CHLRE	Q9fpq6 chlamydomon
11	187	6.2	497	1 WAS2_HUMAN	Q9y6w5 homo sapien
12	185.5	6.1	620	1 EXTN_TOBAC	P13983 nicotiana t
13	182.5	6.0	502	1 WASP_HUMAN	P42768 homo sapien
14	182	6.0	1955	1 PC15_HUMAN	Q96qu1 homo sapien
15	176	5.8	347	1 CSP_PLABA	P23093 plasmodium
16	175.5	5.8	440	1 G3PT_MOUSE	Q64467 mus musculus
17	174.5	5.8	501	1 WASL_RAT	Q08816 rattus norv
18	174.5	5.8	1790	1 SEPA_EMENI	P78621 emericellia
19	174	5.8	431	1 ACRO_RABIT	P48038 oryctolagus
20	174	5.8	520	1 WASP_MOUSE	P70315 mus musculus
21	171.5	5.7	487	1 EBN2_EBV	P12978 Epstein-bar
22	171.5	5.7	1059	1 CAPU_DROME	Q24120 drosophila
23	170.5	5.6	505	1 WASL_HUMAN	Q00401 homo sapien
24	167.5	5.5	339	1 CSP_PLABE	P06915 plasmodium
25	167.5	5.5	1248	1 DIAL_MOUSE	Q06010 homo sapien
26	167.5	5.5	2715	1 TRX2_HUMAN	Q9um66 homo sapien
27	166.5	5.5	1822	1 ZAP3_HUMAN	P49750 homo sapien
28	166.5	5.5	3063	1 CA1C_HUMAN	Q95715 homo sapien
29	165.5	5.5	505	1 WASL_BOVIN	Q95107 bos taurus
30	162	5.4	559	1 WASL_HUMAN	Q92558 homo sapien
31	161.5	5.3	224	1 Y091_NPVAC	P41479 autographa
32	161.5	5.3	421	1 ACRO_HUMAN	P10323 homo sapien
33	160.5	5.3	426	1 EXLP_TOBAC	Q03211 nicotiana t

34	160.5	5.3	1255	1 DIAL_MOUSE	O08808 mus musculus
35	160.5	5.3	1386	1 ZAP3_MOUSE	Q9r017 mus musculus
36	159.5	5.3	503	1 WAIP_HUMAN	O43516 homo sapien
37	159.5	5.3	1162	1 ITAD_HUMAN	Q13349 homo sapien
38	158.5	5.2	449	1 APG_BRANA	P40603 brassica na
39	158.5	5.2	3421	1 TEGU_HSVB	P28955 equine herp
40	158	5.2	261	1 PRP2_MOUSE	P05142 mus musculus
41	157.5	5.2	267	1 EXTN_MAIZE	P14918 zea mays (m
42	157.5	5.2	534	1 APG_ARATH	P40602 arabidopsis
43	157.5	5.2	815	1 MK07_HUMAN	Q13164 homo sapien
44	157	5.2	635	1 VP40_HSV11	P10210 herpes simp
45	157	5.2	819	1 FVB_MOUSE	O35601 mus musculus

## ALIGNMENTS

RESULT 1					
ID	ATR_HUMAN	STANDARD;	PRT;	564 AA.	
AC	Q9H6X2; Q9NVP3;				
DT	01-MAR-2002 (Rel. 41, Created)				
DT	01-MAR-2002 (Rel. 41, Last sequence update)				
DT	01-MAR-2002 (Rel. 41, Last annotation update)				
DE	Anthrax toxin receptor precursor (Tumor endothelial marker 8).				
GN	ATR OR TEM8.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ISOFORM 1).				
RX	MEDLINE=20407466; PubMed=10947988;				
RA	St Croix B., Rago C., Velculescu V., Traverso G., Romans K.E.,				
RA	Montgomery E., Lal A., Riggs G.J., Lengauer C., Vogelstein B.,				
RA	Kinzler K.W.;				
RT	"Genes expressed in human tumor endothelium.";				
RL	Science 289:1197-1202(2000).				
RN	[2]				
RP	SEQUENCE FROM N.A. (ISOFORM 2).				
RX	PubMed=11700562;				
RA	Bradley K.A., Mogridge J., Mourez M., Collier R.J., Young J.A.T.;				
RT	"Identification of the cellular receptor for anthrax toxin.";				
RL	Nature 414:225-229(2001).				
RN	[3]				
RP	SEQUENCE OF 184-564 FROM N.A. (ISOFORM 1), AND SEQUENCE FROM N.A.				
RP	(ISOFORM 3).				
RA	Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,				
RA	Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,				
RA	Tanaka T., Nakamura Y., Isogai T., Sugano S.;				
RT	"NEDO human cDNA sequencing project.";				
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.				
RN	[4]				
RP	SEQUENCE FROM N.A. (ISOFORM 4).				
RA	Traustberg R.;				
RT	TISSUE=Kidney;				
RL	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.				
CC	-!- FUNCTION: Cellular role is not yet known. (PA) of Bacillus				
CC	anthracis.				
CC	-!- SUBUNIT: Binds to the protective antigen (PA) of Bacillus				
CC	anthracis.				
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).				
CC	-!- ALTERNATIVE PRODUCTS: 4 isoforms: 1 (shown here), 2, 3 and 4; seem				
CC	to be produced by alternative splicing.				
CC	-!- TISSUE SPECIFICITY: Highly expressed in tumor endothelial cells				
CC	but not in normal endothelial cells.				
CC	-!- DOMAIN: Binding to PA seems to be effected through the VWA domain.				
CC	-!- SIMILARITY: BELONGS TO THE ATR FAMILY.				
CC	-!- SIMILARITY: CONTAINS 1 VWEA DOMAIN.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AF279145; AAK52094.1; -  
 CC EMBL; AK025429; BAB15128.1; ALT\_INIT.  
 CC EMBL; AK001463; BAA91707.1; ALT\_FRAME.  
 CC EMBL; BC012074; AAH12074.1; -  
 CC MIM; 606410; -  
 CC InterPro; IPR002035; VWFA.  
 CC SMART; SM00327; VWFA; 1.  
 CC PROSITE; PS0234; VWFA; 1.  
 CC Receptor; Transmembrane; Glycoprotein; Signal; Alternative splicing.  
 CC SIGNAL 1 32 POTENTIAL.  
 CC CHAIN 33 564 ANTRAX TOXIN RECEPTOR.  
 CC DOMAIN 33 321 EXTRACELLULAR (POTENTIAL).  
 CC TRANSMEM 322 342 POTENTIAL.  
 CC DOMAIN 343 564 CYTOPLASMIC (POTENTIAL).  
 CC DOMAIN 44 215 VWFA.  
 CC DOMAIN 360 368 ASP/GLU-RICH (HIGHLY ACIDIC).  
 CC DOMAIN 506 564 PRO-RICH.  
 CC CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC VARSPLIC 365 368 EDDD -> NKIK (IN ISOFORM 2).  
 CC VARSPLIC 369 564 MISSING (IN ISOFORM 2).  
 CC VARSPLIC 268 297 NKPFSVEDYLLCPAPILKEVGMKAALQV -> SKSLQSP  
 CC WNTSGFGKEGSHPCPLPARPHT (IN ISOFORM 3).  
 CC MISSING (IN ISOFORM 3).  
 CC DGSILAIALILFELL -> LHKIASGPTTAACME (IN  
 CC ISOFORM 4).  
 CC MISSING (IN ISOFORM 4).  
 CC VARSPLIC 334 564 MISSING (IN ISOFORM 4).  
 CC VARSPLIC 319 333 DGSILAIALILFELL -> LHKIASGPTTAACME (IN  
 CC ISOFORM 4).  
 CC VARSPLIC 334 564 MISSING (IN ISOFORM 4).  
 CC SEQUENCE 564 AA; 62789 MW; B1I8A00AD5DF2233 CRC64;

Query Match 100.0%; Score 3025; DB 1; Length 564;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-176;  
 Matches 564; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATAERRALGIGFQWLSLATLVLCAGOGGRDGGPACYGFDLYFLDKSGSVLHWN 60  
 DB 1 MATAERRALGIGFQWLSLATLVLCAGOGGRDGGPACYGFDLYFLDKSGSVLHWN 60

QY 61 EYIYFVEQLAHKFIQPOLRMSFIVSTGRTTLMKLTEDREQIRGLEELQKVLPGGDTYM 120  
 DB 61 EYIYFVEQLAHKFIQPOLRMSFIVSTGRTTLMKLTEDREQIRGLEELQKVLPGGDTYM 120

QY 121 HEGFERASEQIYYENRQGYRTASVIALTDGELHEDLFFYSREANRSDLGAIYVCV 180  
 DB 121 HEGFERASEQIYYENRQGYRTASVIALTDGELHEDLFFYSREANRSDLGAIYVCV 180

QY 181 KDNETQLARIADSKDHVPVNDGFOALOGIHSILKSCIELAAEPSTICAGESFOVY 240  
 DB 181 KDNETQLARIADSKDHVPVNDGFOALOGIHSILKSCIELAAEPSTICAGESFOVY 240

QY 241 VRNGFRHARNVDRVLCSEKINDSVTLNKPFSVEDYLLCPAPILKEVGMKAALQVSN 300  
 DB 241 VRNGFRHARNVDRVLCSEKINDSVTLNKPFSVEDYLLCPAPILKEVGMKAALQVSN 300

QY 301 DGLSFTSSSVIITTHCSGSIILAIALLILFLLALLLWFWPLCTVIKEVPPPAE 360  
 DB 301 DGLSFTSSSVIITTHCSGSIILAIALLILFLLALLLWFWPLCTVIKEVPPPAE 360

QY 361 ESEEDDDGLPKKWPVDASYGGRGVGGIKRMEYRWGKSTEGAKLEKAKNARVKM 420  
 DB 361 ESEEDDDGLPKKWPVDASYGGRGVGGIKRMEYRWGKSTEGAKLEKAKNARVKM 420

QY 421 PEQYEFEPERNLNNMRRSSPKWYSPKIGKLDALVLLRKYDQVSVNRPQDGTGR 480  
 DB 421 PEQYEFEPERNLNNMRRSSPKWYSPKIGKLDALVLLRKYDQVSVNRPQDGTGR 480

QY 481 CINETRVKNNOPAKYPLNNAYHTSSPPAPITYTPPPAPHCPPPPSPAPTPPSPSTL 540  
 DB 481 CINETRVKNNOPAKYPLNNAYHTSSPPAPITYTPPPAPHCPPPPSPAPTPPSPSTL 540

Db 481 CINETRVKNNOPAKYPLNNAYHTSSPPAPITYTPPPAPHCPPPPSPAPTPPSPSTL 540

Qy 541 PPPQAPPNNAPPPSRPPRPSV 564  
 |||||  
 Db 541 PPPQAPPNNAPPPSRPPRPSV 564

RESULT 2  
 ATR\_MOUSE STANDARD; PRT; 562 AA.  
 ID ATR\_MOUSE  
 AC Q9CZ52;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Antrax toxin receptor precursor (Tumor endothelial marker 8).  
 DE ATR OR TEM8.  
 GN Mus musculus (Mouse).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=21443268; PubMed=11559528;  
 RA Carson-Walter E.B., Watkins D.N., Nanda A., Vogelstein B.,  
 RA Kinzler K.W., St Croix B.;  
 RT "Cell surface tumor endothelial markers are conserved in mice and  
 RT humans.";  
 RL Cancer Res. 61:6649-6655(2001).  
 RN [2]  
 RP SEQUENCE OF 88-562 FROM N.A. (ISOFORM 2).  
 RC MEDLINE=C57BL/6J; TISSUE=Embryo;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,  
 RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Suzuki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 CC -!- FUNCTION: Cellular role is not yet known.  
 CC -!- SUBUNIT: Binds to the protective antigen (PA) of Bacillus  
 CC anthracis (by similarity).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).  
 CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; seem to  
 CC be produced by alternative splicing.  
 CC -!- DOMAIN: Binding to PA seems to be effected through the VWA domain  
 CC (by similarity).  
 CC -!- SIMILARITY: BELONGS TO THE ATR FAMILY.  
 CC -!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AF378762; AAL11999.1; -  
 DR EMBL; AK013005; BAB28591.1; ALT\_INIT.



Fri Aug 9 10:57:06 2002

QY 423 QEVEFFEPRLNNMRRPSSPR-----KWSPKIGKLDLWLLRKGVDVSMRQP 474  
 Db 318 ETEETIRP-----RPPRPKTHPPQTKWYTPKIGRLDALWALLRQYDRVSLMRQP 369  
 QY 475 PGD 477  
 Db 370 EGD 372

RESULT 4  
 CAIC\_CHICK STANDARD; PRT; 3124 AA.  
 AC PI3944; Q04509;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 1(XII) chain precursor (Fibrochimerin).  
 GN COL12A1.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OC NCBI\_TaxID=9031;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WHITE LEGHORN;  
 RX MEDLINE=92011862; PubMed=1918137;  
 RA Yanagata M., Yamada K.M., Yamada S.S., Shinomura T., Tanaka H.,  
 RA Nishida Y., Obara M., Kimata K.;  
 RT "The complete primary structure of type XII collagen shows a chimeric  
 RT molecule with reiterated fibronectin type III motifs, von Willebrand  
 RT factor A motifs, a domain homologous to a noncollagenous region of  
 RT type IX collagen, and short collagenous domains with an Arg-Gly-Asp  
 RT site.";  
 RL J. Cell Biol. 115:209-221(1991).  
 RN [2]  
 RP SEQUENCE OF 2456-3124 FROM N.A., AND SEQUENCE OF 2772-2794 AND  
 RP 2846-2873.  
 RX MEDLINE=90062079; PubMed=2584192;  
 RA Gordon M.K., Gerecke D.R., Dublet B., van der Rest M., Olsen B.R.;  
 RT "Type XII collagen. A large multidomain molecule with partial  
 RT homology to type IX collagen.";  
 RL J. Biol. Chem. 264:19772-19778(1989).  
 RN [3]  
 RP SEQUENCE OF 2960-3076 FROM N.A.  
 RX MEDLINE=87317590; PubMed=3476525;  
 RA Gordon M.K., Gerecke D.R., Olsen B.R.;  
 RT "Type XII collagen: distinct extracellular matrix component  
 RT discovered by cDNA cloning.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:6040-6044(1987).  
 RN [4]  
 RP SEQUENCE OF 1-1283 FROM N.A. (SHORT FORM), AND ALTERNATIVE SPLICING.  
 RC TISSUE=Embryo;  
 RA Trueb J., Trueb B.;  
 RL "The two splice variants of collagen XII share a common 5' end.";  
 RL Biochim. Biophys. Acta 1171:97-98(1992).  
 RN [5]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE=95370352; PubMed=7642694;  
 RA Koch M., Bohrmann B., Matthison M., Hagios C., Trueb B., Chiquet M.;  
 RT "Large and small splice variants of collagen XII: differential  
 RT expression and ligand binding.";  
 RL J. Cell Biol. 130:1005-1014(1995).  
 CC -1- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-  
 CC CONTAINING FIBRILS, THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE  
 CC SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE  
 CC LOCALIZED IN THE PERIFIBRILLAR MATRIX.  
 CC -1- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 kDa OF  
 CC NONTRIPLE-HELICAL SEQUENCES.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE FINAL TISSUE

FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OF EITHER THE LONGER  
 OR THE SHORTER ISOFORM OR ANY COMBINATION OF LONG AND SHORT  
 ISOFORM CHAINS. ONLY THE LONG VARIANT IS A PROTEOGLYCAN. THE LARGE  
 ISOFORM HAS MORE RESTRICTED EXPRESSION IN EMBRYONIC TISSUE THAN  
 THE SMALL.  
 -1- TISSUE SPECIFICITY: TYPE XII COLLAGEN IS PRESENT IN TENDONS,  
 LIGAMENTS, PERICHONDRUM, AND PERIOSTEUM, ALL DENSE CONNECTIVE  
 TISSUES CONTAINING TYPE I COLLAGEN.  
 -1- DOMAIN: THIS SEQUENCE DEFINES FIVE DISTINCT DOMAINS, TWO TRIPLE-  
 HELICAL DOMAINS (COL1 AND COL2) AND THREE NONTRIPLE-HELICAL  
 DOMAINS (NC1, NC2, AND NC3).  
 -1- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT  
 EACH END.  
 -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 -1- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY).  
 -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH  
 INTERRUPTED HELICES (FACIT) FAMILY.  
 -1- SIMILARITY: CONTAINS 4 VWFA DOMAINS.  
 -1- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; D00824; BAA00701.1; -;  
 CC EMBL; X61024; CAA43358.1; -;  
 CC EMBL; M17375; AAA48718.1; -;  
 CC EMBL; J05137; AAA48635.1; -;  
 CC EMBL; X67327; CAA47744.1; -;  
 CC PIR; A28037; A28037.  
 CC PIR; A34485; A34485.  
 CC HSSP; P17301; LAOX.  
 CC InterPro; IPR000087; Collagen.  
 CC InterPro; IPR003961; FN\_III.  
 CC InterPro; IPR003962; FNIII\_repeat.  
 CC InterPro; IPR003129; TSPN.  
 CC InterPro; IPR002035; VWFA.  
 CC Pfam; PF01391; Collagen; 4.  
 CC Pfam; PF00041; fn3; 17.  
 CC Pfam; PF02210; TSPN; 1.  
 CC Pfam; PF00092; vwa; 4.  
 CC PRINTS; PR00014; FNTYPEIII.  
 CC PRINTS; PR00453; VWFADOMAIN.  
 CC SMART; SM00060; FN3; 17.  
 CC SMART; SM00210; TSPN; 1.  
 CC SMART; SM00327; VWA; 4.  
 CC PROSITE; PS50234; VWFA; 4.  
 CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 CC Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.  
 KW SIGNAL  
 KW CHAIN 1 24  
 FT CHAIN 25 3124  
 FT DOMAIN 25 114  
 FT DOMAIN 139 311  
 FT DOMAIN 332 425  
 FT DOMAIN 439 615  
 FT DOMAIN 629 720  
 FT DOMAIN 721 811  
 FT DOMAIN 812 904  
 FT DOMAIN 905 998  
 FT DOMAIN 999 1085  
 FT DOMAIN 1086 1178  
 FT DOMAIN 1199 1371  
 FT DOMAIN 1384 1473  
 FT DOMAIN 1474 1565  
 FT DOMAIN 1566 1654  
 FT DOMAIN 1655 1755  
 FT DOMAIN 1756 1846  
 FT DOMAIN 1847 1936



Fri Aug 9 10:57:06 2002

```

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC -----
CC EMBL; X02873; CAA26632.1; -
CC EMBL; M11221; AAA33137.1; -
CC DR PIR; A24354; A24354.
CC KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
CC FT SIGNAL 1 32
CC FT CHAIN 33 306 EXTENSIN.
CC SQ SEQUENCE 306 AA; 34225 MW; AEC944007F0938DD CRC64;
-----
Query Match 6.3%; Score 191.5; DB 1; Length 306;
Best Local Similarity 26.1%; Pred. No. 5.3e-05;
Matches 59; Conservative 18; Mismatches 56; Indels 93; Gaps 10;
CC
CC 352 KEVPPP---PAESEEEDDDGLPKKKWPTVDASYGGRGVGKIKRMEVRWGEKSTEEGA 408
CC Db 144 KSPPPPKHSPAPEHHYKYKSPPPPKHFPAPHHY----- 177
CC
CC 409 KLEKAKNARVKMPEOEYEFPEPRNLNNMRRPSSPRKYSPKIGKLDALWVLLRKYDRV 468
CC QY 178 -KYKSPPPPTPVYKYKSPPP-----PTPVYKSP----- 208
CC Db
CC 469 SYMRPQGTGRCINFTRVKNQPAKYFLNNAYHTSSPPAPLYTPPPAPHPCHPPPP-- 526
CC QY 209 ----PPP-----KHSPA--PVHHYKYKSPPPPTPVYKSPPPPPHSPPPPTPV 249
CC Db
CC 527 ---SAPTPTPIPSPSTLPPPP-----QAPPPNRPAPPP---SRPPRP 561
CC QY 250 YKYKSPPPHHSPP---PTPVYKYKSPPPPMHSPPPPPPVYSPPPPK 292
CC Db
-----
RESULT 8
EM14.MOUSE
ID EM14.MOUSE STANDARD; PRT; 1206 AA.
AC Q05959;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Formin 1 isoform IV (limb deformity protein).
GN FMN OR LD.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=92112033; PubMed=1339380;
RX Grusby-Jackson L., Kuo A., Leder P.;
RT "A variant limb deformity transcript expressed in the embryonic mouse
RT limb defines a novel formin.";
RL Genes Dev. 6:29-37(1992).
CC -1- FUNCTION: IS IMPORTANT IN THE MORPHOGENESIS OF LIMB AND MAY HAVE A
CC FUNCTION IN DIFFERENTIATED CELLS OR BE INVOLVED IN MAINTAINING
CC SPECIFIC DIFFERENTIATED STATES.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS; IA (AC Q05860), IB (AC
CC Q05860), II (AC Q05860), III (AC Q05860) AND IV (SHOWN HERE); ARE
CC PRODUCED BY ALTERNATIVE SPLICING. A VARIATION IN SPLICING IS SEEN
CC AMONG DIFFERENT TISSUES AND DIFFERENT SIZE TRANSCRIPTS EXIST
CC WITHIN ANY ONE TISSUE.
CC -1- TISSUE SPECIFICITY: IT IS FOUND THROUGHOUT THE EMBRYO BUT
CC HAS A FUNCTIONAL ROLE ONLY IN THE KIDNEY AND LIMB.
CC -1- DEVELOPMENTAL STAGE: THIS IS THE ISOFORM FOUND IN THE APICAL
CC ECTODERMAL RIDGE AND THE MESENCHYMAL COMPARTMENT OF THE DEVELOPING
-----
CC
CC 273 PBPSPPPPPPPPPPPPPPPPPPPPPSPSPRKPSPS 307
-----
RESULT 6
YPRO.OWEFU
ID YPRO.OWEFU STANDARD; PRT; 141 AA.
AC P21260; P21261;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical proline-rich protein (Fragment).
OS Oenila fusiformis.
OC Eukaryota; Metazoa; Annelida; Polychaeta; Palpata; Canaliculipata;
OC Sabellida; Oweniidae; Owenia.
OX NCBI_TaxID=6347;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90147742; PubMed=2105723;
RA Bakalaria N., Collet J., Planells R., Thouveny Y., Fontes M.;
RT "Presence in invertebrate genomes of sequences characterized by the
RT repetition of the triplet CcPurine.";
RL Biochem. Biophys. Res. Commun. 166:66-73(1990).
DR PIR; A34043; A34043.
DR PIR; B34043; B34043.
KW Hypothetical protein; DNA-binding.
FT NON_TER 1 1 POLY-PRO.
FT DOMAIN 9 58 H-T-H MOTIF (POTENTIAL).
FT DNA_BIND 98 116
FT NON_TER 141 141
FT SEQUENCE 141 AA; 15745 MW; B294E884D152BDB9 CRC64;
-----
Query Match 6.5%; Score 197.5; DB 1; Length 141;
Best Local Similarity 58.3%; Pred. No. 9.4e-06;
Matches 35; Conservative 2; Mismatches 18; Indels 5; Gaps 2;
CC
CC 502 HTSSPPAPITYTPPPAPHPCHPPPPSPAPTPIPSPSTLPPPPQAPPPNRPAPPSRPPRP 561
CC QY 141 141 141 141 141 141 141 141 141 141 141 141 141 141 141
CC Db 5 HSLTPPPPP---PPPPPP---PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 59
-----
RESULT 7
EXTN_DAUCA
ID EXTN_DAUCA STANDARD; PRT; 306 AA.
AC P06599;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Extensin precursor.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
OC Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
OX NCBI_TaxID=4039;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen J., Varner J.E.;
RT "An extracellular matrix protein in plants: characterization of a
RT genomic clone for carrot extensin.";
RL EMBO J. 4:2145-2151(1985).
CC [2]
RP SEQUENCE OF 264-306 FROM N.A.
RA Chen J., Varner J.E.;
RT "Isolation and characterization of cDNA clones for carrot extensin
RT and a proline-rich 33-kDa protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4399-4403(1985).
CC -1- FUNCTION: STRUCTURAL COMPONENT IN PRIMARY CELL WALL.
CC -1- SUBCELLULAR LOCATION: Extracellular matrix.
CC -1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC GLYCOSYLATED.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-10 IS THE INITIATOR.
-----

```







the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
-----
DR EMBL; AB026542; BA81795.1; -.
DR EMBL; AL022578; CA18609.1; -.
DR EMBL; AF134304; AAD33053.2; -.
DR MIM; 605875; -.
DR InterPro; IPR003124; WH2.
DR Pfam; PF02205; WH2; 1.
DR SMART; SM00246; WH2; 1.
KW Actin-binding.
FT DOMAIN 268 271 POLY-PRO.
FT DOMAIN 318 328 POLY-PRO.
FT DOMAIN 333 338 POLY-PRO.
FT DOMAIN 347 350 POLY-PRO.
FT DOMAIN 364 371 POLY-PRO.
FT DOMAIN 379 382 POLY-PRO.
FT DOMAIN 391 405 POLY-PRO.
FT DOMAIN 435 452 WH2.
FT CONFLICT 1 1 M -> V (IN REF. 3).
FT CONFLICT 64 64 P -> R (IN REF. 1).
FT CONFLICT 74 74 G -> R (IN REF. 1).
FT CONFLICT 93 93 V -> G (IN REF. 1).
FT CONFLICT 144 144 T -> K (IN REF. 1).
FT CONFLICT 162 162 K -> E (IN REF. 1).
FT CONFLICT 169 169 R -> K (IN REF. 1).
FT CONFLICT 206 206 I -> M (IN REF. 1).
FT CONFLICT 221 222 PF -> TS (IN REF. 1).
FT CONFLICT 250 250 S -> P (IN REF. 1).
FT CONFLICT 311 312 SS -> GP (IN REF. 1).
FT CONFLICT 339 339 I -> V (IN REF. 1).
FT CONFLICT 352 352 S -> P (IN REF. 1).
FT CONFLICT 367 367 L -> P (IN REF. 1).
FT CONFLICT 388 388 R -> G (IN REF. 1).
FT CONFLICT 399 399 P -> PP (IN REF. 1).
FT CONFLICT 415 415 V -> I (IN REF. 1).
FT CONFLICT 432 432 I -> V (IN REF. 1).
FT CONFLICT 436 436 H -> R (IN REF. 1).
FT CONFLICT 444 444 C -> R (IN REF. 1).
FT CONFLICT 456 456 Q -> R (IN REF. 1).
FT CONFLICT 461 461 W -> R (IN REF. 1).
FT CONFLICT 492 492 G -> E (IN REF. 1).
FT CONFLICT 497 497 N -> D (IN REF. 1).
SQ SEQUENCE 497 AA; 54094 MW; 5F313599FDF63A6B CRC64;
```

```
Query Match 6.2%; Score 187; DB 1; Length 497;
Best Local Similarity 26.5%; Pred. No. 0.00018;
Matches 75; Conservative 27; Mismatches 89; Indels 92; Gaps 15;

Qy 356 PPPAESEEDDDGLPKKWTVDASY-----GGRGVGIGKRMVWGEKGSTE 405
Dy 130 PPPLNLTPTPRDDGTEALKFVT-DPSYFPLWKKKMLQDTRDKKKKH-RKEKKDNPN 187
Qy 406 EG-----AKLEKAKNARVKKEQYEYEPENLNNNRRSPSPKWSPIKGLDALWV 459
Dy 188 RGNVNPRIKTRKEWEKIKMGQ-----EVESKEKGLPGFGYP--PTLVYQ--NGSIGCV-- 238
Qy 460 LLRKGYDRVSVMRPQDGTGRCINFRVKNQPA-----KYPLNN-----AY 501
Dy 239 ---ENVDASSYPSPQSASPSFSEDLNLPPEAFESFPVDNOROSGLAGPKRSSV 295
Qy 502 HTSSPPPA-PYTP-----PPAPHCP-----PPPSAPT- 530
Dy 296 SPSHPPPAPPLGSPSSKGFAPAPPPPPPMIGIPPPPPPIGFGSGGTGPPPPSSPSF 355
Qy 531 -----PPIPSPPS-----TLPPPPQAPPPNRPSPRPP 560
Dy 356 PPHDFDAAPPPLPPPPPAADYPTLPPPPPLSQPTRGAPPPPPPPPP 398
```

```
RESULT 12
EXTN_TOBAC
ID EXTN_TOBAC STANDARD; PRT; 620 AA.
AC PI3983;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Extensin precursor (Cell wall hydroxyproline-rich glycoprotein).
DE HKGPNW3.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. XANTHI; TISSUE=Leaf;
RX MEDLINE=90128263; PubMed=2612909;
RA Keller B., Lamb C.J.;
RT "Specific expression of a novel cell wall hydroxyproline-rich
glycoprotein gene in lateral root initiation.";
RL Genes Dev. 3:1639-1646(1989).
CC -!- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
MAIN ROOT.
CC -!- SUBCELLULAR LOCATION: Extracellular matrix.
CC -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
GLYCOSYLATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).

```
-----
EMBL; X13885; CAA32090.1; -.
DR PIR; S06733; S06733.
KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
KW Hydroxylation.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 620 EXTENSIN.
FT REPEAT 70 73 H-A-P-P.
FT REPEAT 148 151 H-A-P-P.
FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
FT REPEAT 229 235 1.
FT REPEAT 236 242 2.
FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.
FT DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.
SQ SEQUENCE 620 AA; 65406 MW; 641DD2278AB28524 CRC64;
```



Query Match 6.1%; Score 185.5; DB 1; Length 620;  
Best Local Similarity 29.7%; Pred. No. 0.00028;  
Matches 52; Conservative 10; Mismatches 34; Indels 79; Gaps 9;



Qy 428 PEPRNLNNRRSPSPKRWYSPIKGLDALWLLRKGYDRVSVMRPQDGTGRCINFRV 487  
Dy 254 PSPP-----RQPPPT--YSP-----PPPA-----YAQS 275  
Qy 488 KNNQAKYPLNNAYHTSSPPAPIYTPPPPA-----PHCPP 524  
Dy 276 PQPSPTSPSPPTTY--SPPPSPSYSPSPSPPTPTFTSPSPSPSPSPPT 333  
Qy 525 PPS-----APTPI--PSPSTLPPPP-----QAPPNRPAPPSPRPPRS 563  
Dy 334 PPTYLPSPSPSYSPSPSPSPSPSPSPSPPTTYLPSPSPSPSPSPSPSPPT 388



RESULT 13


```

Fri Aug 9 10:57:06 2002

```

WASP_HUMAN STANDARD; PRT; 502 AA.
ID P42768; Q9UNJ9;
AC 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Wiskott-Aldrich syndrome protein (WASP).
DE WAS OR IMD2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE="T-cell";
RX MEDLINE=94349367; PubMed=8069912;
RA Derry J.M.J., Ochs H.D., Francke U.;
RT "Isolation of a novel gene mutated in Wiskott-Aldrich syndrome.";
RL Cell 78:635-644(1994).
RN [2]
RP ERATUM.
RX MEDLINE=95094263; PubMed=8001129;
RA Derry J.M.J., Ochs H.D., Francke U.;
RL Cell 79:923-923(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX Kwan S.P., Hagemann T.L., Radtke B.E., Blaese R.M., Rosen F.S.;
RA "Identification of mutations in the Wiskott-Aldrich syndrome gene and
RT characterization of a polymorphic dinucleotide repeat at DXS6940,
RT adjacent to the disease gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:4706-4710(1995).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=99167346; PubMed=10066431;
RA Hagemann T.L., Kwan S.-P.;
RT "The identification and characterization of two promoters and the
RT complete genomic sequence for the Wiskott-Aldrich syndrome gene.";
RL Biochem. Biophys. Res. Commun. 256:104-109(1999).
RN [5]
RP SEQUENCE FROM N.A.
RA Blechschmidt K., Nyakatura G., Strom T.M., Drescher B., Menzel U.,
RA Meindl A., Rosenthal A.;
RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
RN [6]
RP VARIANTS WAS K-31; M-75; P-82; C-86; H-86; C-97; K-133; E-476 AND
RP H-30 DEL.
RX MEDLINE=96133285; PubMed=8528198;
RA Kolluri R., Shehabeldin A., Peacocke M., Lamhonwah A.-M.,
RA Teichert-Kuliszewska K., Weissman S.M., Siminovich K.A.;
RT "Identification of WASP mutations in patients with Wiskott-Aldrich
RT syndrome and isolated thrombocytopenia reveals allelic heterogeneity
RT at the WAS locus.";
RL Hum. Mol. Genet. 4:1119-1126(1995).
RN [7]
RP VARIANTS WAS F-27; I-48; M-75; L-86; H-86; K-131; C-187 AND K-477.
RX MEDLINE=96133286; PubMed=8528199;
RA Derry J.M.J., Kerns J.A., Weinberg K.I., Ochs H.D., Volpini V.,
RA Estivill X., Walker A.P., Francke U.;
RT "WASP gene mutations in Wiskott-Aldrich syndrome and X-linked
RT thrombocytopenia.";
RL Hum. Mol. Genet. 4:1127-1135(1995).
RN [8]
RP VARIANTS WAS VAL-56 AND GLU-236.
RX MEDLINE=9531593; PubMed=7795648;
RA Villa A., Notarangelo L., Macchi P., Mantuano E., Cavagni G.,
RA Brugnani D., Strina D., Patrosso M.C., Ramenghi U., Sacco M.G.,
RA Ugazio A., Vezzoni P.;
RT "X-linked thrombocytopenia and Wiskott-Aldrich syndrome are allelic
RT diseases with mutations in the WASP gene.";
RL Nat. Genet. 9:414-417(1995).
RN [9]
RP VARIANTS WAS TRP-43; MET-45; MET-75 AND CYS-86.
RX
RA
RA Remold-O'Donnell E., Cooley J., Shcherbina A., Hagemann T.L.,
RA Kwan S.-P., Kenney D.M., Rosen F.S.;
RT "Variable expression of WASP in B cell lines of Wiskott-Aldrich
RT syndrome patients.";
RL J. Immunol. 158:4021-4025(1997).
RN [10]
RP VARIANTS WAS LYS-31 AND MET-45.
RX MEDLINE=97253413; PubMed=9098856;
RA Ariga T., Yamada M., Sakiyama Y.;
RT "Mutation analysis of five Japanese families with Wiskott-Aldrich
RT syndrome and determination of the family members' carrier status
RT using three different methods.";
RL Pediatr. Res. 41:535-540(1997).
RN [11]
RP VARIANTS WAS MET-75; LEU-84; ASP-89 AND LYS-133.
RX MEDLINE=98350091; PubMed=9683546;
RA MacCarthy-Morrogh L., Gaspar H.B., Wang Y.-C., Thompson L.,
RA Layton M., Jones A.M., Kinnon C.;
RT "Absence of expression of the Wiskott-Aldrich syndrome protein in
RT peripheral blood cells of Wiskott-Aldrich syndrome patients.";
RL Clin. Immunol. Immunopathol. 88:22-27(1998).
RN [12]
RP VARIANTS WAS VAL-56.
RX MEDLINE=98378988; PubMed=9713366;
RA Facchetti F., Blanzuoli L., Vermi W., Notarangelo L.D., Gilliani S.,
RA Fiorini M., Fasth A., Stewart D.M., Nelson D.L.;
RT "Defective actin polymerization in EBV-transformed B-cell lines from
RT patients with the Wiskott-Aldrich syndrome.";
RL J. Pathol. 185:99-107(1998).
RN [13]
RP VARIANTS WAS LYS-133.
RX MEDLINE=98092074; PubMed=9445409;
RA Parolini O., Ressmann G., Haas O.A., Pawlowsky J., Gadner H.,
RA Knapp W., Holter W.;
RT "X-linked Wiskott-Aldrich syndrome in a girl.";
RL New Engl. J. Med. 338:291-295(1998).
RN [14]
RP VARIANTS WAS ARG-73; MET-75; CYS-83; CYS-86 AND LYS-133.
RX MEDLINE=99374925; PubMed=10447259;
RA Lemahieu V., Gastier J.M., Francke U.;
RT "Novel mutations in the Wiskott-Aldrich syndrome protein gene and
RT their effects on transcriptional, translational, and clinical
RT phenotypes.";
RL Hum. Mutat. 14:54-66(1999).
RN [15]
RP VARIANTS WAS HIS-52 AND TRP-70.
RX MEDLINE=11793485;
RA El-Hakeh J., Rosenzweig S., Oleastro M., Basack N., Berozdnic L.,
RA Molina F., Rivas E.M., Zelazko M., Danielian S.;
RT "Wiskott-Aldrich syndrome in Argentina: 17 unique, including nine
RT novel, mutations.";
RL Hum. Mutat. 19:186-187(2002).
RN [16]
RP FUNCTION: POSSIBLE REGULATOR OF LYMPHOCYTE AND PLATELET FUNCTION.
CC MAY BE INVOLVED IN SIGNALING PATHWAYS WITH CYTOSKELETAL FUNCTION.
CC
CC -1- SUBUNIT: BINDS TO CDC42, RAC, RCK, FYN, SRC KINASE FGR, BTK, ABL,
CC WIP, AND TO THE P85 SUBUNIT OF PLC-GAMMA.
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE THYMUS. ALSO
CC FOUND, TO A MUCH LESSER EXTENT, IN THE SPLEEN.
CC -1- DOMAIN: THE WH1 (WASP HOMOLGY 1) DOMAIN MAY BIND A PRO-RICH
CC LIGAND.
CC -1- DISEASE: DEFECTS IN WAS ARE THE CAUSE OF WISKOTT-ALDRICH SYNDROME
CC (WAS), AN X-LINKED RECESSIVE IMMUNODEFICIENCY CHARACTERIZED BY
CC ECZEMA, THROMBOCYTOPENIA, RECURRENT INFECTIONS, AND BLOODY
CC DIARRHEA. DEATH USUALLY OCCURS BEFORE AGE 10.
CC -1- DISEASE: DEFECTS IN WAS ARE THE CAUSE OF ISOLATED X-LINKED
CC THROMBOCYTOPENIA (XLT). XLT IS CLINICALLY MILD WITH SMALL
CC PLATELETS AND SUBCLINICAL LEUKOCYTE ABNORMALITIES.
CC -1- SIMILARITY: CONTAINS 1 GBD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 WH1 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation

```



```

OX NCBI_TaxID=5823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-90221834; PubMed-2183186;
RA Lockyer M.J., Davies C.S., Suhrbier A., Sinden R.E.:
RT "Nucleotide sequence of the Plasmodium berghei circumsporozoite
RL protein gene from the ANKA clone 2.34L.";
RL Nucleic Acids Res. 18:376-376(1990).
CC -!- FUNCTION: THE CIRCUSSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE
CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC VERTEBRATE HOST).
CC -!- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
CC -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X17606; CAA35608.1; -.
DR PIR; S07873; OZZQBK.
DR InterPro; IPR003067; Crmsprzoite.
DR InterPro; IPR000884; TSPl.
DR Pfam; PF00090; tsp_1; 1.
DR PRINTS; PR01303; CRCMSRZRTOE.
DR SMART; SM00209; TSPI; 1.
DR PROSITE; PS00092; TSPI; 1.
KW Malaria; Sporozoite; Repeat; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 347 CIRCUSSPOROZOITE PROTEIN.
FT DOMAIN 93 204 13 X 8 AA REPEATS.
FT DOMAIN 214 247 17 X 2 AA REPEATS OF P-Q.
FT SEQUENCE 347 AA; 37776 MW; 0EC240EE35681AF8 CRC64;
SQ

Query Match 5.8%; Score 176; DB 1; Length 347;
Best Local Similarity 34.6%; Pred. No. 0.00054;
Matches 47; Conservative 14; Mismatches 55; Indels 20; Gaps 7;

QY 431 RLNLNNMRPSSPRKWSIPKGLDALWLLRKGYDRVSVMPQCDTGRCINFTRVKNN 490
| | | : : : : : | : | : | : | : |
DB 32 RLNLNELCYNEGDNKLHYLVNSKNKGKIYT--RNTVNRLLADAPE-----SK-----KNE 78
| | | : : : : : | : | : | : | : |

QY 491 QPAKYPLNNAYHTSSPPAPIVTPPPPAPHCPHP--PSAPTPTPIPSPPTLTLPFPPQPAPP 549
| | | | | | | | | | | | | | | | | | |
DB 79 KKKNIERNKCLKOPPPNP-NDPPPNNDPPPNPDPPPNPDPP--PePNNDPPP 135
| | | | | | | | | | | | | | | | | | |

QY 550 ---NRAPPPSRPPRP 562
| | | | | | | | | | | | | | | | | | |
DB 136 PNANDPPPNANDPAP 151
| | | | | | | | | | | | | | | | | | |

Search completed: August 9, 2002, 10:47:04
Job time: 827 sec
```

Search completed: August 9, 2002, 10:47:04  
Job time: 827 sec

..

Fri Aug 9 10:57:06 2002

us-09-970-076-6.rsp

Page 13

**THIS PAGE BLANK (18PTO)**  
**THIS PAGE BLANK (18PTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:47:04 ; Search time 31.3 seconds  
(without alignments)  
411.936 Million cell updates/sec

Title: US-09-970-076-8

Perfect score: 1728

Sequence: 1 MATAERRALGIGFQWLSLAT.....TTHCSLHKIASGPTTAACME 333

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	1649	95.4	564	1 ATR_HUMAN	Q9h6x2 homo sapien
2	1353	89.9	562	1 ATR_MOUSE	Q9cz52 mus musculus
3	551.5	31.9	386	1 CMG2_HUMAN	P58335 homo sapien
4	159.5	9.2	1162	1 ITAD_HUMAN	Q13349 homo sapien
5	145.5	8.4	1152	1 ITAM_HUMAN	P11215 homo sapien
6	143.5	8.3	1163	1 ITAX_HUMAN	P20702 homo sapien
7	143	8.3	3124	1 CAIC_CHICK	P13944 gallus gall
8	141	8.2	1153	1 ITAM_MOUSE	P05555 mus musculus
9	139	8.0	3067	1 CAIC_MOUSE	Q60847 mus musculus
10	131	7.6	760	1 CO2_MOUSE	P21180 mus musculus
11	131	7.6	3063	1 CAIC_HUMAN	Q99715 homo sapien
12	130	7.5	1888	1 CALE_CHICK	P32018 gallus gall
13	128	7.4	3051	1 YN33_CAEEL	P34576 caenorhabdi
14	120.5	7.0	496	1 CANA_HUMAN	P21941 homo sapien
15	118	6.8	929	1 CAIC_NOTVI	Q91145 notophthalm
16	117.5	6.8	452	1 MTN3_CHICK	Q42401 gallus gall
17	114	6.6	956	1 MTN2_HUMAN	Q00339 homo sapien
18	113.5	6.6	764	1 CPAB_HUMAN	P00751 homo sapien
19	111.5	6.5	2944	1 CA17_HUMAN	Q02388 homo sapien
20	110.5	6.4	761	1 CFAB_MOUSE	P04186 mus musculus
21	109.5	6.3	1029	1 CA26_MOUSE	Q02788 mus musculus
22	108.5	6.3	500	1 CANA_MOUSE	P15942 mus musculus
23	108.5	6.3	624	1 MTN4_MOUSE	O89029 mus musculus
24	108.5	6.3	956	1 MTN2_MOUSE	O08746 mus musculus
25	108	6.2	1189	1 ITAH_HUMAN	Q9ukx5 homo sapien
26	107.5	6.2	493	1 CANA_CHICK	P05099 gallus gall
27	103	6.0	619	1 MTN3_HUMAN	O95460 homo sapien
28	102.5	5.9	486	1 MTN3_HUMAN	O15232 homo sapien
29	102	5.9	639	1 CAIC_RABIT	Q28902 cryotolagus
30	99.5	5.8	597	1 ROP_DROME	Q07327 drosophila
31	99	5.7	341	1 Y173_BORBU	O51195 borrelia bu
32	99	5.7	559	1 TRAP_PLAFA	P16893 plasmodium
33	99	5.7	1018	1 CA26_HUMAN	P12110 homo sapien

#### RESULT 1

ID	ATR_HUMAN	STANDARD;	PRT;	564 AA.
AC	Q9H6X2: Q9NVP3;			
DT	01-MAR-2002 (Rel. 41, Created)			
DT	01-MAR-2002 (Rel. 41, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Anthrax toxin receptor precursor (Tumor endothelial marker 8).			
GN	ATR OR TEM8.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RX	MEDLINE=20407466; PubMed=10947988;			
RA	St Croix B., Rago C., Velculescu V., Traverso G., Romans K.E.,			
RA	Montgomery E., Lal A., Riggins G.J., Lengauer C., Vogelstein B.,			
RA	Kinzler K.W.;			
RT	"Genes expressed in human tumor endothelium.";			
RL	Science 289:1197-1202(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A. (ISOFORM 2).			
RX	PubMed=11700562;			
RA	Bradley K.A., Mogridge J., Mourez M., Collier R.J., Young J.A.T.;			
RT	"Identification of the cellular receptor for anthrax toxin.";			
RL	Nature 414:225-229(2001).			
RN	[3]			
RP	SEQUENCE OF 184-564 FROM N.A. (ISOFORM 1), AND SEQUENCE FROM N.A.			
RP	(ISOFORM 3).			
RA	Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,			
RA	Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,			
RA	Tanaka T., Nakamura Y., Isogai T., Sugano S.;			
RT	"NEDO human cDNA sequencing project.";			
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A. (ISOFORM 4).			
RP	TISSUE=Kidney;			
RA	Strausberg R.;			
RT	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.			
CC	!- FUNCTION: Cellular role is not yet known.			
CC	!- SUBUNIT: Binds to the protective antigen (PA) of Bacillus anthracis.			
CC	!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).			
CC	!- ALTERNATIVE PRODUCTS: 4 isoforms; 1 (shown here), 2, 3 and 4; seem to be produced by alternative splicing.			
CC	!- TISSUE SPECIFICITY: Highly expressed in tumor endothelial cells but not in normal endothelial cells.			
CC	!- DOMAIN: Binding to PA seems to be effected through the VWA domain.			
CC	!- SIMILARITY: BELONGS TO THE ATR FAMILY.			
CC	!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way			

34	98	5.7	3176	1	CA36_HUMAN	P12111 homo sapien
35	97.5	5.6	401	1	HUTI_CAOCR	P58079 caulobacter
36	97.5	5.6	642	1	DNAK_NEIMB	O9K0N4 neisseria m
37	97	5.6	285	1	ITAL_CHICK	O90615 gallus gall
38	97	5.6	3137	1	CA36_CHICK	P19989 gallus gall
39	95.5	5.5	1179	1	ITAE_HUMAN	P38570 homo sapien
40	95	5.5	1167	1	ITAE_MOUSE	Q60677 mus musculus
41	94.5	5.5	1151	1	ITAL_HUMAN	P56199 homo sapien
42	94	5.4	1062	1	NAL2_HUMAN	Q9nx02 homo sapien
43	94	5.4	1170	1	ITP2_BOVIN	P53710 bos taurus
44	93	5.4	1180	1	ITAL_RAT	P18614 ratus norv
45	92.5	5.4	550	1	COCH_HUMAN	O43405 homo sapien

#### ALIGNMENTS





```
FT HCPPDPAPSAPTPIPPSPSTLPPPPQAPPNNRAPPSPRPPP
FT RSV -> RFRGWRITICLSKSHVHGRDKGPETPLKQA
FT WMFSFLERAFO (IN ISOFORM 2).
SQ SEQUENCE 562 AA; 62308 MW; 6AC92049B4BB4F7C CRC64;

Query Match      89.9%; Score 1553; DB 1; Length 562;
Best Local Similarity 96.1%; Pred. No. 2.2e-120;
Matches 298; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 9 LGIGFOWLSLATLVLCACGGRRDGGPACYGDFLYFLDKSGSVLHHWNEIYFVQ 68
DB 7 LGAGRLGCVAAVLVCAAGHRRDGGPACYGDFLYFLDKSGSVLHHWNEIYFVQ 66

QY 69 LAHKFISPOLRMSFVFSRGTTLTKMLTDEQIROGLELEOKLPGGDTYMHGFERAS 128
DB 67 LAHRFISPOLRMSFVFSRGTTLTKMLTDEQIROGLELEOKLPGGDTYMHGFERAS 126

QY 129 EQIYYENROGYRTASVIIALTDLGELHEDLFFYSERANRSDLGAIYVCVGVKDFNETOL 188
DB 127 EQIYYENROGYRTASVIIALTDLGELHEDLFFYSERANRSDLGAIYVCVGVKDFNETOL 186

QY 189 ARTADSKDHVPVNDGFQALQGIHHSILKSCIEILAAEPSTICAGESQVYVVRNGFRH 248
DB 187 ARTADSKDHVPVNDGFQALQGIHHSILKSCIEILAAEPSTICAGESQVYVVRNGFRH 246

QY 249 ARNVDRVLCSEKINDSVTNEKPFVEDTYLLCPALKEVGMKAQLOVSMNDGLSFISS 308
DB 247 ARNVDRVLCSEKINDSVTNEKPFVEDTYLLCPALKEVGMKAQLOVSMNDGLSFISS 306

QY 309 SVIITTHCS 318
DB 307 SVIITTHCS 316

RESULT 3
CMG2_HUMAN STANDARD; PRT; 386 AA.
AC P58335;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Capillary morphogenesis protein-2 precursor (CMG-2).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21539596; PubMed=11683410;
RA Bell S.E., Mavila A., Salazar R., Bayless K.J., Kanagala S.,
RA Maxwell S.A., Davis G.E.;
RT "Differential gene expression during capillary morphogenesis in 3D
RT collagen matrices: regulated expression of genes involved in basement
RT membrane matrix assembly, cell cycle progression, cellular
RT differentiation and G-protein signaling.";
RL J. Cell Sci. 114:2755-2773(2001).
CC 1- SUBUNIT: Seems to bind to collagen type IV and laminin.
CC 1- SUBCELLULAR LOCATION: type I membrane protein (probable).
CC 1- SIMILARITY: BELONGS TO THE ATR FAMILY.
CC 1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AV040326; AAK77222.1;
DR PROSITE; PS50234; VWFA; 1.
KW Transmembrane; Glycoprotein; Signal.
```

```
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 386 CAPILLARY MORPHOGENESIS PROTEIN-2.
FT DOMAIN 34 215 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 216 236 POTENTIAL.
FT DOMAIN 237 386 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 44 213 VWFA.
FT CARBOHYD 211 211 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 386 AA; 42918 MW; 0F9B2D688EAB17A CRC64;

Query Match      31.9%; Score 551.5; DB 1; Length 386;
Best Local Similarity 53.8%; Pred. No. 4e-38;
Matches 114; Conservative 38; Mismatches 57; Indels 3; Gaps 2;

QY 4 AERRALGIGFOWLSLATLVLCACGGG-RREDGGPACYGDFLYFLDKSGSVLHHWNEI 62
DB 3 AERSPARSGSWLPGCLWLVLSGPGGLLRAQEPSCRAFDLYFLDKSGSVANNWIEI 62

QY 63 YFVEQLAHKFISPOLRMSFVFSRGTTLTKMLTDEQIROGLELEOKLPGGDTYMHME 122
DB 63 YNFVQOLAERFVSPENRLSFVFSQAATILPLTGDGRKISKGLDLKRVSPVGETYIHE 122

QY 123 GPERASEQIYYENROGYRTASVIIALTDLGELHEDLFFYSERANRSDLGAIYVCVGVKD 182
DB 123 GLKLANEQI--OKRAGLKTSSIIIALTDGKLDGLVSPYAEKEAKISRSLGASVYCVGVLD 180

QY 183 FNETOLARIADSKDHVPVNDGFQALQGIHHS 214
DB 181 FEQAQLERIADSKDHVPVNDGFQALQGIHHS 212

RESULT 4
ITAD_HUMAN STANDARD; PRT; 1162 AA.
AC Q13349; Q15575; Q15576;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-D precursor (Leukointegrin alpha D) (CD11d) (ADB2).
GN ITGAD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Splicein;
RX MEDLINE=961111956; PubMed=8777714;
RA Van der Vliet M., Le Trong H., Wood C.L., Moore P.F., St John T.,
RA Staunton D.E., Gallatin W.M.;
RT "A novel leukointegrin, alpha d beta 2, binds preferentially to ICAM-
RT 3.";
RL Immunity 3:683-690(1995).
RN [2]
RP SEQUENCE OF 1-235 FROM N.A.
RX MEDLINE=20187620; PubMed=10722744;
RA Noti J.D., Johnson A.K., Dillon J.D.;
RT "Structural and functional characterization of the leukocyte integrin
RT gene CD11d. Essential role of Spi and Sp3.";
RL J. Biol. Chem. 275:8959-8969(2000).
RN [3]
RP SEQUENCE OF 467-537; 571-602; 633-666; 788-834 AND 910-1125 FROM N.A.
RX MEDLINE=96257236; PubMed=8666289;
RA Wong D.A., Davis E.M., LeBeau M., Springer T.A.;
RT "Cloning and chromosomal localization of a novel gene encoding a human
RT beta 2-integrin alpha subunit.";
RL Gene 171:291-294(1996).
RN [4]
RP INTERACTION WITH VCAM1.
RX MEDLINE=99059842; PubMed=9841932;
RA Grayson M.H., Van der Vliet M., Sterbinsky S.A., Michael Gallatin W.,
RA Hoffman P.A., Staunton D.E., Bochner B.S.;
RT "alphadbeta2 integrin is expressed on human eosinophils and functions
```

as an alternative ligand for vascular cell adhesion molecule 1  
(VCAM-1).; J. Exp. Med. 188:2187-2191(1998).

[5]

INTERACTION WITH VCAM1.  
MEDLINE=9370002; PubMed=10438935;  
Van der Vliet M., Crowe D.T., Hoekstra D., Vazeux R., Hoffman P.A.,  
Grayson M.H., Bochner B.S., Gallatin W.M., Staunton D.E.;  
"The leukocyte integrin alpha D beta 2 binds VCAM-1: evidence for a  
binding interface between I domain and VCAM-1.";  
J. Immunol. 163:1984-1990(1999).

FUNCTION: INTEGRIN ALPHA-D/BETA-2 IS A RECEPTOR FOR ICAM3 AND  
VCAM1. MAY PLAY A ROLE IN THE ATHEROSCLEROTIC PROCESS SUCH AS  
CLEARING LIPOPROTEINS FROM PLAQUES AND IN PHAGOCYTOSIS OF BLOOD-

BORNE PATHOGENS, PARTICULATE MATTER, AND SENESECENT ERYTHROCYTES

FROM THE BLOOD.

SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-D

ASSOCIATES WITH BETA-2.

SUBCELLULAR LOCATION: Type I membrane protein.

TISSUE SPECIFICITY: EXPRESSED MODERATELY ON MYELOMONOCYTIC CELL

LINE AND SUBSETS OF PERIPHERAL BLOOD LEUKOCYTES AND STRONGLY ON

TISSUE-SPECIALIZED CELLS, INCLUDING MACROPHAGES FOAM CELLS WITHIN

ATHEROSCLEROTIC PLAQUES, AND ON SPLENIC RED PULP MACROPHAGES.

DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS

WITH I-DOMAINS DO NOT UNDERGO PROTEINASE CLEAVAGE. FAMILY.

SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.

SIMILARITY: CONTAINS 1 VWFA DOMAIN.

SIMILARITY: CONTAINS 7 FG-GAP REPEATS.

-----  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U37028; AAB38547.1; -

EMBL; U40274; AAB60634.1; -

EMBL; U40275; AAB60635.1; -

EMBL; U40276; AAB60636.1; -

EMBL; U40277; AAB60637.1; -

EMBL; U40279; AAB60638.1; -

EMBL; U40278; AAB60638.1; JOINED.

EMBL; AF187881; AAF62875.1; -

HSP; P11215; IABX.

MIM; 602453; -

InterPro; IPR000413; Integrin\_alpha.

InterPro; IPR002035; VWFA.

Pfam; PF01839; FG-GAP; 5.

Pfam; PF00357; integrin\_A; 1.

Pfam; PF00092; vwa; 1.

PRINTS; PR01185; INTEGRINA.

PRINTS; PR00453; VWFADOMAIN.

SMART; SM00191; Int\_alpha; 4.

SMART; SM00327; VWA; 1.

PROSITE; PS00242; INTEGRIN\_ALPHA; 1.

PROSITE; PS50234; VWFA; 1.

Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;

Signal; Repeat; Calcium;

Magnesium.

SIGNAL

CHAIN

DOMAIN

DOMAIN

TRANSMEM

DOMAIN

REPEAT

REPEAT

REPEAT

REPEAT

REPEAT

REPEAT

REPEAT

REPEAT

REPEAT

REPEAT

REPEAT	581	633	FG-GAP 7.
CA_BIND	465	473	POTENTIAL.
CA_BIND	530	538	POTENTIAL.
CA_BIND	593	601	POTENTIAL.
SITE	1127	1131	GFGR MOTIF.
DISULFID	67	74	BY SIMILARITY.
DISULFID	106	124	BY SIMILARITY.
DISULFID	655	710	BY SIMILARITY.
DISULFID	769	775	BY SIMILARITY.
DISULFID	846	861	BY SIMILARITY.
DISULFID	994	1018	BY SIMILARITY.
DISULFID	1023	1028	BY SIMILARITY.
CARBOHYD	59	59	N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD	87	87	N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD	99	99	N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD	391	391	N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD	691	691	N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD	733	733	N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD	873	873	N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD	957	957	N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD	1046	1046	N-LINKED (GLCNAC. .) (POTENTIAL).
CONFLICT	500	500	MISSING (IN REF. 2).
CONFLICT	515	518	GHEW -> ATP (IN REF. 2).
CONFLICT	825	825	L -> V (IN REF. 2).
CONFLICT	984	984	V -> A (IN REF. 2).
SEQUENCE	1162 AA;	126885 MW;	F296AIA3545D77D CRC64;

Query Match 9.2% Score 159.5; DB 1; Length 1162;

Best Local Similarity 23.4% Pred. No. 3.1e-05;

Matches 70; Conservative 54; Mismatches 124; Indels 51; Gaps 12;

QY 11 IGFWLSLTLVLCAGGRRGGPAC-YGGFDLFLDKSGSV-LHHWNEIYFVEQ 68

DB 127 LGSREIITQV-----PDATPCPHQEMDVLFDGSGSDONDFNQMKFVQA 175

QY 69 LAHFISPLQRMSFTVSTRTTLMLKLTEDREQLEELQKVLV-----GGDTYMEHGF 124

DB 176 VMGQEGTDTLTFALMOYS-----NLLKIHFTFTQFTSPSQSLVDPIVLKGLTFTATGI 231

QY 125 ERASEQIYYENRQGYRTA-SVITALTGELHEDLFFYSERANSRDLGAIVCVGVKDF 183

DB 232 LTVVTQLFHHKNGARKSAKILIVITDGQYKDKPLEYSD-VIPOAKAGIIRYAIQVCHA 290

QY 184 NETQLAR-----TADSKDHVPVNDQFOALQGIHISILKKSCTEILAAEPSTICAGES 236

DB 291 FQGPARTARQELNTISSAPPQDHVKV-DNFAALGSIQKQLEK-----IVAVEGTQSRASS 345

QY 237 FQVVVRGNGFRHARNVDRLVC-----SFKINDS-----VTLEKPFSEDTYL 279

DB 346 FQHEMSQEGFSTALTMGDLGLGAVGFSWGGAFLYPPNMSPTFFINMSQENVMDRDSYL 404

RESULT 5

ITAM\_HUMAN

ITAM\_HUMAN

AC P11215;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Integrin alpha-M precursor (Cell surface glycoprotein MAC-1 alpha

subunit) (CR-3 alpha chain) (CD11b) (Leukocyte adhesion receptor M01)

DE (Neutrophil adherence receptor).

GN ITGAM OR CR3A OR CD11B.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=88315033; PubMed=2457584;

RA Corbi A.L., Kishimoto T.K., Miller L.J., Springer T.A.;

RT "The human leukocyte adhesion glycoprotein Mac-1 (complement receptor

RT type 3, CD11b) alpha subunit. Cloning, primary structure, and  
RL relation to the integrins, von Willebrand factor and factor B.";  
RN J. Biol. Chem. 263:12403-12411(1988).  
RP [2]  
RX SEQUENCE FROM N.A.  
RA MEDLINE=88190151; PubMed=2833753;  
RA Arnaout M.A., Remold-O'Donnell E., Pierce M.W., Harris P., Tenen D.G.;  
RT "Molecular cloning of the alpha subunit of human and guinea pig  
RT leukocyte adhesion glycoprotein Mol: chromosomal localization and  
RL homology to the alpha subunits of integrins.";  
RN Proc. Natl. Acad. Sci. U.S.A. 85:2776-2780(1988).  
RP [3]  
RX SEQUENCE FROM N.A.  
RA MEDLINE=88257215; PubMed=2454931;  
RA Arnaout M.A., Gupta S.K., Pierce M.W., Tenen D.G.;  
RT "Amino acid sequence of the alpha subunit of human leukocyte adhesion  
RT receptor Mol (complement receptor type 3).";  
RL J. Cell Biol. 106:2153-2158(1988).  
RP [4]  
RX SEQUENCE FROM N.A.  
RA MEDLINE=93123748; PubMed=8419480;  
RA Fleming J.C., Pahl H.L., Gonzalez D.A., Smith T.F., Tenen D.G.;  
RT "Structural analysis of the CD11b gene and phylogenetic analysis of  
RT the alpha-integrin gene family demonstrate remarkable conservation of  
RT genomic organization and suggest early diversification during  
RL evolution.";  
RN J. Immunol. 150:480-490(1993).  
RP [5]  
RX SEQUENCE OF 1-9 FROM N.A.  
RA MEDLINE=9098893; PubMed=2563162;  
RA Hickstein D.D., Hickey M.J., Ozols J., Baker D.M., Back A.L.,  
RA Roth G.J.;  
RT "cDNA sequence for the alpha M subunit of the human neutrophil  
RT adherence receptor indicates homology to integrin alpha subunits.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:257-261(1989).  
RP [6]  
RX SEQUENCE OF 1-9 FROM N.A.  
RA MEDLINE=92073318; PubMed=1683702;  
RA Shelley C.S., Arnaout M.A.;  
RT "The promoter of the CD11b gene directs myeloid-specific and  
RT developmentally regulated expression.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:10525-10529(1991).  
RP [7]  
RX SEQUENCE OF 1-9 FROM N.A.  
RA TISSUE=Blood;  
RA MEDLINE=92144986; PubMed=1346576;  
RA Pahl H.L., Rosmarin A.G., Tenen D.G.;  
RT "Characterization of the myeloid-specific CD11b promoter.";  
RL Blood 79:865-870(1992).  
RP [8]  
RX SEQUENCE OF 17-31.  
RA MEDLINE=87076671; PubMed=3539202;  
RA Pierce M.W., Remold-O'Donnell E., Todd R.F. III, Arnaout M.A.;  
RT "N-terminal sequence of human leukocyte glycoprotein Mol:  
RT conservation across species and homology to platelet IIb/IIIa.";  
RL Biochim. Biophys. Acta 874:368-371(1986).  
RP [9]  
RX X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 148-331.  
RA MEDLINE=95171458; PubMed=7867070;  
RA Lee J.O., Rieu P., Arnaout M.A., Liddington R.;  
RT "Crystal structure of the A domain from the alpha subunit of integrin  
RT CR3 (CD11b/CD18).";  
RL Cell 80:631-638(1995).  
RP [10]  
RX X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 148-334.  
RA MEDLINE=96363671; PubMed=8747460;  
RA Lee J.O., Bankston L.A., Arnaout M.A., Liddington R.C.;  
RT "Two conformations of the integrin A-domain (I-domain): a pathway for  
RT activation?";  
RL Structure 3:1333-1340(1995).  
RP [11]  
RX X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 148-337.  
RA MEDLINE=98362595; PubMed=9687375;

RA Baldwin E.T., Sarver R.W., Bryant G.L. Jr., Curry K.A.,  
RA Fairbanks M.B., Finzel B.C., Garlick R.L., Heinrikson R.L.,  
RA Horton N.C., Kelley L.L., Mildner A.M., Moon J.B., Mott J.E.,  
RA Mutschler V.T., Tomich C.S., Watenpaugh K.D., Wiley V.H.;  
RT "Cation binding to the integrin CD11b I domain and activation model  
RT assessment.";  
RL Structure 6:923-935(1998).  
RP [12]  
RX 3D-STRUCTURE MODELING OF 17-616.  
RA MEDLINE=98226734; PubMed=9560195;  
RA Oxvig C., Springer T.A.;  
RT "Experimental support for a beta-propeller domain in integrin alpha-  
RT subunits and a calcium binding site on its lower surface.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:4870-4875(1998).  
RP [13]  
RX FUNCTION: INTEGRIN ALPHA-M/BETA-2 IS IMPLICATED IN VARIOUS  
RA ADHESIVE INTERACTIONS OF MONOCYTES, MACROPHAGES AND GRANULOCYTES  
CC AS WELL AS IN MEDIATING THE UPTAKE OF COMPLEMENT-COATED PARTICLES.  
CC THE THIRD COMPLEMENT COMPONENT. IT PROBABLY RECOGNIZES THE R-G-D  
CC PEPTIDE IN C3B. INTEGRIN ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR  
CC FIBRINOGEN, FACTOR X AND ICAM1. IT RECOGNIZES P1 AND P2 PEPTIDES  
CC OF FIBRINOGEN GAMMA CHAIN.  
CC [14]  
RX SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-M  
CC ASSOCIATES WITH BETA-2.  
CC [15]  
RX SUBCELLULAR LOCATION: Type I membrane protein.  
CC [16]  
RX TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND  
CC GRANULOCYTES.  
CC [17]  
RX DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS  
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE. FAMILY.  
CC [18]  
RX SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
CC [19]  
RX SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
CC [20]  
RX SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
CC [21]  
RX DATABASE: NAME=PRO; NOTE=CD guide CD11b entry;  
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd11b.htm".  
CC [22]  
RX This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC [23]  
RX EMBL; J03925; AAA59544.1; -  
DR EMBL; M18044; AAA59491.1; -  
DR EMBL; J04145; AAA59503.1; -  
DR EMBL; S52227; AAB24821.1; -  
DR EMBL; S52152; AAB24821.1; JOINED.  
DR EMBL; S52153; AAB24821.1; JOINED.  
DR EMBL; S52154; AAB24821.1; JOINED.  
DR EMBL; S52155; AAB24821.1; JOINED.  
DR EMBL; S52157; AAB24821.1; JOINED.  
DR EMBL; S52159; AAB24821.1; JOINED.  
DR EMBL; S52161; AAB24821.1; JOINED.  
DR EMBL; S52164; AAB24821.1; JOINED.  
DR EMBL; S52165; AAB24821.1; JOINED.  
DR EMBL; S52167; AAB24821.1; JOINED.  
DR EMBL; S52169; AAB24821.1; JOINED.  
DR EMBL; S52170; AAB24821.1; JOINED.  
DR EMBL; S52173; AAB24821.1; JOINED.  
DR EMBL; S52174; AAB24821.1; JOINED.  
DR EMBL; S52180; AAB24821.1; JOINED.  
DR EMBL; S52181; AAB24821.1; JOINED.  
DR EMBL; S52184; AAB24821.1; JOINED.  
DR EMBL; S52189; AAB24821.1; JOINED.  
DR EMBL; S52191; AAB24821.1; JOINED.  
DR EMBL; S52192; AAB24821.1; JOINED.  
DR EMBL; S52203; AAB24821.1; JOINED.  
DR EMBL; S52212; AAB24821.1; JOINED.  
DR EMBL; S52213; AAB24821.1; JOINED.  
DR EMBL; S52216; AAB24821.1; JOINED.  
DR EMBL; S52219; AAB24821.1; JOINED.  
DR EMBL; S52220; AAB24821.1; JOINED.





This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see [http://www.isb-sib.ch/announcement/send\\_email\\_to\\_license@isb-sib.ch](http://www.isb-sib.ch/announcement/send_email_to_license@isb-sib.ch)).

R	EMBL; D00824; BRA00701.1; -	
R	EMBL; X61024; CAA4358.1; -	
R	EMBL; M17375; AAA48718.1; -	
R	EMBL; J05137; AAA48635.1; -	
R	EMBL; X67327; CAA47744.1; -	
R	PIR; A28037; A28037.	
R	PIR; A34485; A34485.	
R	HSSP; P17301; IA0X.	
R	InterPro; IPR000087; Collagen.	
R	InterPro; IPR003961; FN.III.	
R	InterPro; IPR003962; FN.II_repeat.	
R	InterPro; IPR003129; TSPN.	
R	InterPro; IPR002035; VWFA.	
R	Pfam; PF01391; Collagen; 4.	
R	Pfam; PF00041; fn3; 17.	
R	Pfam; PF02210; TSPN; 1.	
R	Pfam; PF00092; vwa; 4.	
R	PRINTS; PR00014; FNTYPE.III.	
R	PRINTS; PR00453; VWFADOMAIN.	
R	SMART; SM00060; FN3; 17.	
R	SMART; SM00210; TSPN; 1.	
R	SMART; SM00327; VWA; 4.	
R	PROSITE; PS50234; VWFA; 4.	
R	W	Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
R	W	Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.
T	SIGNAL	1 24 POTENTIAL.
T	CHAIN	25 3124 COLLAGEN ALPHA 1(XII) CHAIN.
T	DOMAIN	25 114 FIBRONECTIN TYPE-III 1.
T	DOMAIN	139 311 VWFA 1.
T	DOMAIN	332 425 FIBRONECTIN TYPE-III 2.
T	DOMAIN	439 615 VWFA 2.
T	DOMAIN	629 720 FIBRONECTIN TYPE-III 3.
T	DOMAIN	721 811 FIBRONECTIN TYPE-III 4.
T	DOMAIN	812 904 FIBRONECTIN TYPE-III 5.
T	DOMAIN	905 998 FIBRONECTIN TYPE-III 6.
T	DOMAIN	999 1085 FIBRONECTIN TYPE-III 7.
T	DOMAIN	1086 1178 FIBRONECTIN TYPE-III 8.
T	DOMAIN	1199 1371 VWFA 3.
T	DOMAIN	1384 1473 FIBRONECTIN TYPE-III 9.
T	DOMAIN	1474 1565 FIBRONECTIN TYPE-III 10.
T	DOMAIN	1566 1654 FIBRONECTIN TYPE-III 11.
T	DOMAIN	1655 1755 FIBRONECTIN TYPE-III 12.
T	DOMAIN	1756 1846 FIBRONECTIN TYPE-III 13.
T	DOMAIN	1847 1936 FIBRONECTIN TYPE-III 14.
T	DOMAIN	1937 2027 FIBRONECTIN TYPE-III 15.
T	DOMAIN	2028 2118 FIBRONECTIN TYPE-III 16.
T	DOMAIN	2119 2206 FIBRONECTIN TYPE-III 17.
T	DOMAIN	2207 2295 FIBRONECTIN TYPE-III 18.
T	DOMAIN	2327 2500 VWFA 4.
T	DOMAIN	2455 2750 NONHELICAL REGION (NC3).
T	DOMAIN	2751 2902 TRIPLE-HELICAL REGION (COL2)
FT	DOMAIN	2903 2945 WITH 1 IMPERFECTION.
FT	DOMAIN	2946 NONHELICAL REGION (NC2).
FT	DOMAIN	3048 TRIPLE-HELICAL REGION (COL1)
FT	DOMAIN	WITH 2 IMPERFECTIONS.
FT	DOMAIN	3049 NONHELICAL REGION (NC1).
FT	DOMAIN	3086 ASP/GLU-RICH (ACIDIC).
FT	DOMAIN	3111 ARG/LYS-RICH (BASIC).
FT	BINDING	797 TO CHONDROITIN SULFATE (POTENTIAL).
FT	BINDING	890 TO CHONDROITIN SULFATE (POTENTIAL).
FT	BINDING	981 TO CHONDROITIN SULFATE (POTENTIAL).
FT	BINDING	2440 CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	2438 CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	2859 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	32 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	1006 1006

FT	CARBOHYD	1032	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1044	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1512	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1767	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2210	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2273	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2532	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2683	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	25	MISSING (IN SHORT ISOFORM).
FT	CONFLICT	1258	T -> S (IN REF. 4).
FT	CONFLICT	1258	D -> E (IN REF. 4).
FT	CONFLICT	1264	P -> A (IN REF. 2).
FT	CONFLICT	2759	L -> F (IN REF. 2).
FT	CONFLICT	2803	V -> F (IN REF. 2).
FT	CONFLICT	2977	QP -> AG (IN REF. 3).
FT	CONFLICT	3076	QP -> AG (IN REF. 3).
FT	CONFLICT	3124	AA; 340578 MW; 094285AFE7F346CF CRC64;
FT	SEQUENCE	3124	AA; 340578 MW; 094285AFE7F346CF CRC64;
QY	Query Match	8.3%;	Score 143; DB 1; Length 3124;
QY	Best Local Similarity	26.1%;	Pred. No. 0.0025;
QY	Matches	65;	Conservative 51; Mismatches 97; Indels 36; Gaps
QY	44	DLXFLDKSGSV-LHWHNEIYFVEQLAHKF-ISP-OLRMSFIVFSTRGTTLMKLKTEDE	100
QY	439	DVWFLVDGYSIGIANFVKVRAFLVLYLVSFEISPRKVOISLVQISR--DPHMEFSLNRY	496
QY	101	QIRQGLEELQKVLPP--GGDFYMHGEPERASEQIYYENRQGYR--TASVIAIATDGGELHED	156
QY	497	NRVKDIIQAINTPYRGSGTNTGKATYYREKVFVTSK--GSRENVPVRMILITDGK--SSD	554
QY	157	LFYFSEARENRSDLGAIVYCVGKDFNETQLARIAD--SKDHVPVNDGFOALQGLIHS	214
QY	555	AF---KEPAIKLRDADVEIFAVGKDAVRTELEAIASPPAEHTHYTVED-FDAFQRISE	610
QY	215	ILKKSCTEIIAAEPSTICAGESQVYVVGNGFRHARNV---DRLVCSKFINDSVTLNEKP	271
QY	611	LQOSVCLRI-----EQELAAIRKSKYVPKNNVFSDDTSDFKVSNSAAGSE--	657
QY	272	FSVEDTVLL	280
QY	658	---EKSVI	663
RESULT	8		
ITAM_MOUSE	STANDARD;	PRT;	1153 AA.
ID	ITAM_MOUSE		
AC	P0555;		
DT	01-NOV-1988 (Rel. 09, Created)		
DT	01-FEB-1991 (Rel. 17, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Integrin alpha-M precursor (Cell surface glycoprotein MAC-1 alpha subunit) (CR-3 alpha chain) (CD11b) (Leukocyte adhesion receptor MO1)		
GN	ITGAM.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE=88312584; PubMed=3044779;		
RA	Pytela R.;		
RT	"Amino acid sequence of the murine Mac-1 alpha chain reveals homology with the integrin family and an additional domain related to von Willebrand factor."		
RT	Willembrand factor."		
RL	EMBO J. 7:1371-1378(1988).		
RN	[2]		
RP	SEQUENCE OF 11-45 FROM N.A.		
RC	STRAIN=BALB/C; TISSUE=Spleen;		
RX	MEDLINE=86287312; PubMed=2942940;		
RA	Sastre L., Roman J.M., Teplov D.B., Dreyer W.J., Gee C.E.,		
RA	Larson R.S., Roberts T.M., Springer T.A.;		
RT	"A partial genomic DNA clone for the alpha subunit of the mouse complement receptor type 3 and cellular adhesion molecule Mac-1."		
RT	complement receptor type 3 and cellular adhesion molecule Mac-1."		

RL Proc. Natl. Acad. Sci. U.S.A. 83:5644-5648(1986).  
 RN [3]  
 RP SEQUENCE OF 17-28.  
 RX MEDLINE=8518276; PubMed=3987182;  
 RA Springer T.A., Teplov D.B., Dreyer W.J.;  
 RT "Sequence homology of the LFA-1 and Mac-1 leukocyte adhesion  
 RT glycoproteins and unexpected relation to leukocyte interferon.";  
 RL Nature 314:540-542(1985).  
 CC -1- FUNCTION: INTEGRIN ALPHA-M/BETA-2 IS IMPLICATED IN VARIOUS  
 CC ADHESIVE INTERACTIONS OF MONOCYTES, MACROPHAGES AND GRANULOCYTES  
 CC AS WELL AS IN MEDIATING THE UPTAKE OF COMPLEMENT-COATED PARTICLES.  
 CC IT IS IDENTICAL WITH CR-3, THE RECEPTOR FOR THE IC3B FRAGMENT OF  
 CC THE THIRD COMPLEMENT COMPONENT. IT PROBABLY RECOGNIZES THE R-G-D  
 CC PEPTIDE IN C3B. INTEGRIN ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR  
 CC FIBRINOGEN, FACTOR X AND ICAM1. IT RECOGNIZES P1 AND P2 PEPTIDES  
 CC OF FIBRINOGEN GAMMA CHAIN. ALPHA-M/BETA-2 PLAY A CRITICAL ROLE IN  
 CC MAST CELL DEVELOPMENT AND IN IMMUNE COMPLEX-MEDIATED  
 CC GLOMERULONEPHRITIS. MICE EXPRESSING A NULL MUTATION OF THE ALPHA-M  
 CC SUBUNIT GENE DEMONSTRATE INCREASE IN NEUTROPHIL ACCUMULATION, IN  
 CC RESPONSE TO A IMPAIRED DEGRANULATION AND PHAGOCYTOSIS, EVENTS THAT  
 CC APPARENTLY ACCELERATE APOPTOSIS IN NEUTROPHILS. THESE MICE DEVELOP  
 CC OBESITY.  
 CC -1- SUBUNIT: HETERO-DIMER OF AN ALPHA AND A BETA CHAIN. ALPHA-M  
 CC ASSOCIATES WITH BETA-2.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND  
 CC GRANULOCYTES  
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS  
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X07640; CAA30479.1; -;  
 DR EMBL; M14293; AAA39484.1; -;  
 DR PIR; S00551; S00551.  
 DR HSP; P11215; IABX.  
 DR MGD; MGI:96607; Itgam.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; VWFA.  
 DR Pfam; PF01839; FG-GAP; 5.  
 DR Pfam; PF00357; Integrin\_A; 1.  
 DR Pfam; PF00092; vwa; 1.  
 DR PRINTS; PR01185; INTEGRINA.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR SMART; SM00191; Int\_alpha; 4.  
 DR SMART; SM00327; VWFA; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS0234; VWFA; 1.  
 KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
 KW Signal; Calcium; Repeat.  
 FT SIGNAL 1 16  
 FT CHAIN 17 1153  
 FT DOMAIN 17 1105  
 FT TRANSMEM 1106 1129  
 FT DOMAIN 1130 1153  
 FT REPEAT 31 84  
 FT REPEAT ? ?  
 FT DOMAIN 164 350  
 FT REPEAT 337 400  
 FT REPEAT 401 452  
 FT REPEAT 454 515  
 FT REPEAT 517 575  
 FT REPEAT 580 632  
 FT REPEAT 655 743  
 FT CA\_BIND 465 473  
 FT POTENTIAL.

FT CA\_BIND 529 537 POTENTIAL.  
 FT CA\_BIND 592 600 POTENTIAL.  
 FT SITE 1132 1136 GFFKR MOTIF.  
 FT DISULFID 66 73 BY SIMILARITY.  
 FT DISULFID 105 123 BY SIMILARITY.  
 FT DISULFID 654 711 BY SIMILARITY.  
 FT DISULFID 770 776 BY SIMILARITY.  
 FT DISULFID 999 1023 BY SIMILARITY.  
 FT DISULFID 1028 1033 BY SIMILARITY.  
 FT CARBOHYD 58 58 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 391 391 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 696 696 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 734 734 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 772 772 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 801 801 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 881 881 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 907 907 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 941 941 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 980 980 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 994 994 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1045 1045 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1051 1051 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1076 1076 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 1153 AA; 127480 MW; 178DB988AECB0343 CRC64;  
 Query Match 8.2%; Score 141; DB 1; Length 1153;  
 Best Local Similarity 24.2%; Pred. No. 0.001;  
 Matches 71; Conservative 50; Mismatches 96; Indels 76; Gaps 16;  
 QY 44 DLYFILDKSGVHLHWNHYIYFVEQLAHKFI---PQLRMSFIVFSTGCTLMKLTED-- 98  
 DB 150 DIVFLIDSGSI-----NNTDF---QMKKEFVSTVMEQFKKSKTLFS-----LMQYSDFR 197  
 QY 99 ----REQIRQGLEELQKVLVLP---GGDTYMHGEGFERASQIYYE-NRQGYRTASVIALT 149  
 DB 198 IHFTFNDFRNPSRSHVSPIKQLNGRTKTASGIRKVVRELFHKTNGARENAKILVIT 257  
 QY 150 DGEHLEDLFYSE--REANRSRDGLAIVYCVGVKD-FNETQLARIAD-----SKDHVFP 200  
 DB 258 DGEKFGDPLDYKDVPEADRA---GVIRYVIGVGNFKNPQSRRELDITASKPAGEHVFQ 314  
 QY 201 VNDGFOALQGIHSLKSCIEFLAAEPCTICAGESFOVVVRGNGFRHARNVDRVLCSPK 260  
 DB 315 V-DNFEALNTIQNLOEK---IFAIEGTGTSTSTSEHEMSQEGF----- 355  
 QY 261 INDSVTLNEKPFESVEDTYLLCPAPILKEVG---MKAALQVSMNDGLSFSS 309  
 DB 356 -SASITSN-----GPLLGSVGSFDWAGGAFLYTSKDKVTFINTT 393  
 RESULT 9  
 ID CA1C\_MOUSE STANDARD; PRT; 3067 AA.  
 AC Q60847.  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 1(XII) chain precursor.  
 GN COL12A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RC STRAIN=SWISS WEBSTER, AND C57BL/6J; TISSUE=Skin;  
 RX MEDLINE=96170761; PubMed=8601036;  
 RA Boehme K., Li Y., Oh P.S., Olsen B.R.;  
 RT "Primary structure of the long and short splice variants of mouse  
 RT collagen XII and their tissue-specific expression during embryonic











FT	DOMAIN	145	192	EGF-LIKE 4.
FT	DOMAIN	194	245	EGF-LIKE 5.
FT	DOMAIN	247	296	EGF-LIKE 6.
FT	DOMAIN	313	354	EGF-LIKE 7.
FT	DOMAIN	315	402	EGF-LIKE 8.
FT	DOMAIN	404	452	EGF-LIKE 9.
FT	DOMAIN	454	503	EGF-LIKE 10.
FT	DOMAIN	514	690	VWFA.
FT	DOMAIN	705	750	EGF-LIKE 11.
FT	DOMAIN	750	794	EGF-LIKE 12.
FT	DOMAIN	805	846	EGF-LIKE 13.
FT	DOMAIN	847	892	EGF-LIKE 14.
FT	DOMAIN	892	940	EGF-LIKE 15.
FT	DOMAIN	942	990	EGF-LIKE 16.
FT	DOMAIN	992	1039	EGF-LIKE 17.
FT	DOMAIN	1043	1091	EGF-LIKE 18.
FT	DOMAIN	1093	1144	EGF-LIKE 19.
FT	DOMAIN	1146	1195	EGF-LIKE 20.
FT	DOMAIN	1197	1245	EGF-LIKE 21.
FT	DOMAIN	1247	1295	EGF-LIKE 22.
FT	DOMAIN	1298	1346	EGF-LIKE 23.
FT	DOMAIN	1352	1396	EGF-LIKE 24.
FT	DOMAIN	1397	1444	EGF-LIKE 25.
FT	DOMAIN	1446	1492	EGF-LIKE 26.
FT	DOMAIN	1494	1542	EGF-LIKE 27.
FT	DOMAIN	1544	1592	EGF-LIKE 28.
FT	DOMAIN	1594	1642	EGF-LIKE 29.
FT	DOMAIN	1644	1692	EGF-LIKE 30.
FT	DOMAIN	1693	1739	EGF-LIKE 31.
FT	DOMAIN	1740	1788	EGF-LIKE 32.
FT	DOMAIN	1797	1847	EGF-LIKE 33.
FT	DOMAIN	1849	1900	EGF-LIKE 34.
FT	DOMAIN	1902	1950	EGF-LIKE 35.
FT	DOMAIN	1952	1998	EGF-LIKE 36.
FT	DOMAIN	2000	2047	EGF-LIKE 37.
FT	DOMAIN	2047	2095	EGF-LIKE 38.
FT	DOMAIN	2117	2156	EGF-LIKE 39.
FT	DOMAIN	2157	2281	SEA 1.
FT	DOMAIN	2293	2332	EGF-LIKE 40.
FT	DOMAIN	2333	2456	SEA 2.
FT	DOMAIN	2460	2504	EGF-LIKE 41.
FT	DOMAIN	2508	2556	EGF-LIKE 42.
FT	DOMAIN	2556	2608	EGF-LIKE 43.
FT	DOMAIN	2612	2657	EGF-LIKE 44.
FT	DOMAIN	2657	2693	EGF-LIKE 45.
FT	DISULFID	2	13	BY SIMILARITY.
FT	DISULFID	7	22	BY SIMILARITY.
FT	DISULFID	24	43	BY SIMILARITY.
FT	DISULFID	50	63	BY SIMILARITY.
FT	DISULFID	57	72	BY SIMILARITY.
FT	DISULFID	74	93	BY SIMILARITY.
FT	DISULFID	104	120	BY SIMILARITY.
FT	DISULFID	112	129	BY SIMILARITY.
FT	DISULFID	131	143	BY SIMILARITY.
FT	DISULFID	149	163	BY SIMILARITY.
FT	DISULFID	157	172	BY SIMILARITY.
FT	DISULFID	174	191	BY SIMILARITY.
FT	DISULFID	198	214	BY SIMILARITY.
FT	DISULFID	208	223	BY SIMILARITY.
FT	DISULFID	225	244	BY SIMILARITY.
FT	DISULFID	251	265	BY SIMILARITY.
FT	DISULFID	259	274	BY SIMILARITY.
FT	DISULFID	276	295	BY SIMILARITY.
FT	DISULFID	317	330	BY SIMILARITY.
FT	DISULFID	324	339	BY SIMILARITY.
FT	DISULFID	341	353	BY SIMILARITY.
FT	DISULFID	359	371	BY SIMILARITY.
FT	DISULFID	365	380	BY SIMILARITY.
FT	DISULFID	382	401	BY SIMILARITY.
FT	DISULFID	408	421	BY SIMILARITY.
FT	DISULFID	415	430	BY SIMILARITY.
FT	DISULFID	432	451	BY SIMILARITY.
FT	DISULFID	458	472	BY SIMILARITY.

```
FT DISULFID 466 481 BY SIMILARITY.
FT DISULFID 483 502 BY SIMILARITY.
FT DISULFID 709 725 BY SIMILARITY.
FT DISULFID 717 734 BY SIMILARITY.
FT DISULFID 736 749 BY SIMILARITY.
FT DISULFID 754 768 BY SIMILARITY.
FT DISULFID 762 778 BY SIMILARITY.
FT DISULFID 780 793 BY SIMILARITY.
FT DISULFID 809 822 BY SIMILARITY.
FT DISULFID 816 831 BY SIMILARITY.
FT DISULFID 833 845 BY SIMILARITY.
FT DISULFID 851 867 BY SIMILARITY.
FT DISULFID 859 876 BY SIMILARITY.
FT DISULFID 878 891 BY SIMILARITY.
FT DISULFID 896 909 BY SIMILARITY.
FT DISULFID 903 918 BY SIMILARITY.
FT DISULFID 920 939 BY SIMILARITY.
FT DISULFID 946 959 BY SIMILARITY.
FT DISULFID 953 968 BY SIMILARITY.
FT DISULFID 970 989 BY SIMILARITY.
FT DISULFID 996 1010 BY SIMILARITY.
FT DISULFID 1004 1019 BY SIMILARITY.
FT DISULFID 1021 1038 BY SIMILARITY.
FT DISULFID 1047 1060 BY SIMILARITY.
FT DISULFID 1054 1070 BY SIMILARITY.
FT DISULFID 1072 1090 BY SIMILARITY.
FT DISULFID 1097 1113 BY SIMILARITY.
FT DISULFID 1105 1122 BY SIMILARITY.
FT DISULFID 1124 1143 BY SIMILARITY.
FT DISULFID 1150 1164 BY SIMILARITY.
FT DISULFID 1157 1173 BY SIMILARITY.
FT DISULFID 1175 1194 BY SIMILARITY.
FT DISULFID 1201 1214 BY SIMILARITY.
FT DISULFID 1208 1223 BY SIMILARITY.
FT DISULFID 1225 1244 BY SIMILARITY.
FT DISULFID 1251 1264 BY SIMILARITY.
FT DISULFID 1258 1273 BY SIMILARITY.
FT DISULFID 1275 1294 BY SIMILARITY.
FT DISULFID 1302 1315 BY SIMILARITY.
FT DISULFID 1309 1324 BY SIMILARITY.
FT DISULFID 1326 1345 BY SIMILARITY.
FT DISULFID 1356 1372 BY SIMILARITY.
FT DISULFID 1364 1381 BY SIMILARITY.
FT DISULFID 1383 1395 BY SIMILARITY.
FT DISULFID 1401 1415 BY SIMILARITY.
FT DISULFID 1409 1424 BY SIMILARITY.
FT DISULFID 1426 1443 BY SIMILARITY.
FT DISULFID 1450 1464 BY SIMILARITY.
FT DISULFID 1458 1473 BY SIMILARITY.
FT DISULFID 1475 1491 BY SIMILARITY.
FT DISULFID 1498 1512 BY SIMILARITY.
FT DISULFID 1506 1521 BY SIMILARITY.
FT DISULFID 1523 1541 BY SIMILARITY.
FT DISULFID 1548 1562 BY SIMILARITY.
FT DISULFID 1556 1571 BY SIMILARITY.
FT DISULFID 1573 1591 BY SIMILARITY.
FT DISULFID 1598 1611 BY SIMILARITY.
FT DISULFID 1605 1620 BY SIMILARITY.
FT DISULFID 1622 1641 BY SIMILARITY.
FT DISULFID 1648 1661 BY SIMILARITY.
FT DISULFID 1655 1670 BY SIMILARITY.
FT DISULFID 1672 1691 BY SIMILARITY.
FT DISULFID 1697 1709 BY SIMILARITY.
FT DISULFID 1703 1719 BY SIMILARITY.
FT DISULFID 1721 1738 BY SIMILARITY.
FT DISULFID 1744 1758 BY SIMILARITY.
FT DISULFID 1752 1767 BY SIMILARITY.
FT DISULFID 1769 1787 BY SIMILARITY.
FT DISULFID 1801 1815 BY SIMILARITY.

Query Match 7.4%; Score 128; DB 1; Length 3051;
Best Local Similarity 27.1%; Pred. No. 0.042;
Matches 64; Conservative 31; Mismatches 75; Indels 66; Gaps 13;

QY 39 CYGGE-----DLVFILDKSGVLHH--WNEIYFVEOLA 70
      |||||
      || ||| ||| ||| |||
Db 483 CYGGEVDVSSNANLPPGRVCTVQTCPKQKTDLVLDGSGSIGSVFKNEVLVRVREFV 542
      |||||
      || ||| ||| ||| ||| |||
QY 71 HKFI--SPOLRMSEFVSTRGTTLMLKTE--DREQIRGCLLELQKVLPG---GDTYMH- 121
      |||||
      || ||| ||| ||| ||| |||
Db 543 ELFEIGRSKTRVGLIQLYSDQIRHFEFDLQYGDGRDLLKGISETQ-YLTGLTGTGAAIQHM 601
      |||||
      || ||| ||| ||| ||| |||
QY 122 --EGFERASEQIYYENROGYR-----TASVIALTDGELHEDLFFYSREANRSDLGAI 174
      |||||
      || ||| ||| ||| ||| |||
Db 602 VQEGF-----SERRGARPOQSDIARVAIILTDGRSQNV-----TCPADSAKRLSIN 648
      |||||
      || ||| ||| ||| ||| |||
QY 175 VYCVGVKD-FNETOLARIADSKDHVPVNDGFOALQGIHHSILKKSCIEILAAEPS 229
      |||||
      || ||| ||| ||| ||| |||
Db 649 TFAIGVTDHVLASELESIAGSPNRWFYV-DKFKDLDTLRLSMIQK-----NACFS 697
      |||||
      || ||| ||| ||| ||| |||

RESULT 14
CAMA_HUMAN
ID CAMA_HUMAN STANDARD; PRT; 496 AA.
AC P21941;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cartilage matrix protein precursor (Matrilin-1).
GN MATN1 OR CRTM OR CMP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RX MEDLINE-91060566; PubMed-2246248;
RA Jenkins R.N., Osborne-Lawrence S.L., Sinclair A.K., Eddy R.L. Jr.,
RA Byers M.G., Shows T.B., Duby A.D.;
RT "Structure and chromosomal location of the human gene encoding
RT cartilage matrix protein.";
RL J. Biol. Chem. 265:19624-19631(1990).
CC -!- FUNCTION: CARTILAGE MATRIX PROTEIN IS A MAJOR COMPONENT OF THE
CC EXTRACELLULAR MATRIX OF NONARTICULAR CARTILAGE. IT BINDS TO
CC COLLAGEN.
CC -!- SUBUNIT: HOMOTRIMER.
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 VWFA DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M55682; AAB38702.1;
CC EMBL; M55675; AAB38702.1; JOINED.
CC EMBL; M55676; AAB38702.1; JOINED.
CC EMBL; M55677; AAB38702.1; JOINED.
CC EMBL; M55679; AAB38702.1; JOINED.
CC EMBL; M55680; AAB38702.1; JOINED.
CC EMBL; M55681; AAB38702.1; JOINED.
CC EMBL; M55683; AAA63904.1; ALT_SEQ.
CC PIR; A37979; A37979.
CC HSP; P05099; IAQ5.
CC MIN; I15437; -.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR002035; VWFA.
CC Pfam; PF00008; EGF; 1.
CC Pfam; PF00092; vwa; 2.
CC PRINTS; PR00453; VWFADOMAIN.
CC SMART; SM00181; EGF; 1.
CC SMART; SM00327; VWA; 2.
CC PROSITE; PS00022; EGF_1; FALSE_NEG.
```

TENDENCY TOWARD DISTAL REGIONALIZATION. CONDENSING CARTILAGE SHOWS NO SIGNAL. FINALLY, AT THE LATE DIGIT STAGE, EXPRESSION BECOMES LARGELY RESTRICTED TO THE PERICHONDRUM.

-1- P1M: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT EACH END (BY SIMILARITY).

-1- P1M: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY SIMILARITY).

-1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH INTERRUPTED HELICES (FACIT) FAMILY.

-1- SIMILARITY: CONTAINS 2 VWFA DOMAINS.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL: U19494; AAA80217.1; -.

CC HSP: P02751; IFNA.

DR InterPro: IPR003961; FN III.

DR InterPro: IPR002035; VWFA.

DR Pfam: PF00041; fn3; 7.

DR Pfam: PF00092; vwa; 1.

DR SMART: SM0060; FN3; 5.

DR SMART: SM00327; VWFA; 1.

DR PROSITE: PS0234; VWFA; 2.

DR Extracellular matrix; Connective tissue; Repeat; Cell adhesion;

KW Collagen; Glycoprotein.

KW NON\_TER 1

FT DOMAIN <1 49

FT DOMAIN 63 154

FT DOMAIN 155 245

FT DOMAIN 246 338

FT DOMAIN 339 432

FT DOMAIN 433 519

FT DOMAIN 520 612

FT DOMAIN 633 805

FT DOMAIN 818 907

FT DOMAIN 908 >929

FT BINDING 231 231

FT BINDING 324 324

FT BINDING 415 415

FT BINDING 98 98

FT CARBOHYD 929 929

FT NON\_TER 929 929

FT SEQUENCE 929 AA; 101647 MW; AED57485254FD954 CRC64;

Query Match 6.8%; Score 118; DB 1; Length 929;

Best Local Similarity 21.2%; Pred. No. 0.061;

Matches 57; Conservative 51; Mismatches 135; Indels 26; Gaps 9;

QY 44 DLYFILDKSGSV-LHHWNEIYFVEQLAHFISPOLRMSFVSTGRTTLMKLTDEQEI 102

DB 633 DIVLLVDGSGISGRPNFIVRNFISRVVEVDIGSDRQVIAVSQYSGDPRTWQLNTHKT 692

QY 103 RQGLEELQKVLV--GGDTYMHGEGFERASEQIYENRQY-----RTASVITALTDELHED 156

DB 693 KKSMDAVANLPYKGGNTGTSALKFLENNF---RPGVGMREKARKIAILLTDGKSQDD 749

QY 157 LFFYSEREANRSLGAIVCVGVKDFNETOLARIADSKHVPFN-DGQALOGIHSI 215

DB 750 IVAPSKRYA---DEGIELYAVGINKADENELKEIASDPDELYMYNVAADFSLLTINVDL 805

QY 216 LKSCIEILAA---EPSTICAGESFQVVRNGNGFRHARNVDRLVCLSKINDSVTLNPKP 271

DB 806 TENVCNSKVGPGGLAPPNSNLVTSFPTPSFRVTVWPPSQSVER----PKVEYYPVAGGRP 861

QY 272 FSVEDTYLCPAPILKEVGMKAALQVSMN 300

DB 862 ---QEVYVGRGTQTTTLVLGLKPKETEVYVN 887

DR PROSITE: PS01186; EGF 2; 1.

KW EGF-like domain; Signal; Glycoprotein; Cartilage; Repeat; Coiled coil.

FT SIGNAL 1 22

FT CHAIN 23 496

FT DOMAIN 23 222

FT DOMAIN 223 263

FT DOMAIN 264 453

FT DOMAIN 467 495

FT CARBOHYD 76 76

FT DISULFID 35 221

FT DISULFID 227 238

FT DISULFID 234 247

FT DISULFID 249 262

FT DISULFID 265 452

FT SEQUENCE 496 AA; 53700 MW; 2D880A8114C7940F CRC64;

Query Match 7.0%; Score 120.5; DB 1; Length 496;

Best Local Similarity 25.3%; Pred. No. 0.017;

Matches 60; Conservative 37; Mismatches 95; Indels 45; Gaps 13;

QY 24 ICAGGGREDDGGPACVGGDFLFDKSGSVL-HHWNEIYFVEQLAHKF-ISPOLRMS 81

DB 264 VCSGGGS-----SATDLVFLDGSKSVRPENFELVKKFISQIVDTLDVSDKLAQV 314

QY 82 FTV---FSTRGTTLMKLTEDREQIRQGLEELQKVLPGDGYMHGEGFERASEQIY-----Y 133

DB 315 GLVQYSSVRQFPLGRFHTKKDKAAVRNM-----SYMKGTTGAALKYLINDSF 366

QY 134 ENRQYRTAS--VITALTDELHEDLFFYSEREANRSLGAIVCVGVKDFNETOLARI 191

DB 367 TVSSGARPGAKGVGVFTDGRSQD-----YINDAKKADLGLKFMFVGVGNVAVDELREI 422

QY 192 ADS--KDHVFPVNDGFOALQGIHSILKSKSGIEILAAEPSTICAGES---FQVVVRG 243

DB 423 ASEPVAEHYFTYAD-FKINQIGKLOKKICVE---EDP---CACESLVKFOAKVEG 472

RESULT 15

ID CAIC NOTVI STANDARD; PRT; 929 AA.

AC Q91145;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Collagen alpha 1(XII) chain (Fragment).

OS Notophthalmus viridescens (Eastern newt) (Triturus viridescens).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Caudata; Salamandroides; Salamandridae;

OC Notophthalmus.

OX NCBI\_TaxID=8316;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95246925; PubMed=7729585;

RA Wei Y., Yang E.V., Klatt K.P., Tassava R.A.;

RT "Monoclonal antibody MT2 identifies the urodele alpha 1 chain of type XII collagen, a developmentally regulated extracellular matrix protein in regenerating newt limbs.";

RT Dev. Biol. 168:503-513(1995).

RL -1- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-CONTAINING FIBRILS, THE COLI DOMAIN COULD BE ASSOCIATED WITH THE SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE LOCALIZED IN THE PERIFIBRILLAR MATRIX (BY SIMILARITY). COULD PLAY A DEVELOPMENTAL ROLE IN REGENERATION.

CC -1- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 KDa OF NONTRIPLE-HELICAL SEQUENCES (BY SIMILARITY).

CC -1- DEVELOPMENTAL STAGE: EXPRESSION STARTS AT 3 DAYS AFTER AMPUTATION IN CELLS OF THE BASAL LAYER OF THE WOUND EPITHELIUM. AT DAY 10, EXPRESSION IS FOUND IN BOTH THE BASAL WOUND EPITHELIAL CELLS AND THE DISTAL MESENCHYME CELLS. AT MID-BUD AND LATE-BUD BLASTOMA STAGES, WOUND EPITHELIAL EXPRESSION HAS DECREASED, WHEREAS THE MESENCHYME REMAINS STRONGLY ACTIVE IN TRANSCRIPTION AND SHOWED A

Search completed: August 9, 2002, 10:47:07  
Job time: 830 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: August 9, 2002, 10:44:32 ; Search time 507.33 Seconds  
(without alignments)  
151.246 Million cell updates/sec

Title: US-09-970-076-10

Perfect score: 1131

Sequence: 1 MSFIVFSTRGTTLMKLTEDR.....STSGFKNGSHPCLPARPH 218

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main: \*  
1: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/paa/US080\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep.\*  
7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep.\*  
8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep.\*  
9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep.\*  
10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep.\*  
11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep.\*  
12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep.\*  
13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep.\*  
14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep.\*  
15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep.\*  
16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep.\*  
17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep.\*  
18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep.\*  
19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep.\*  
20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep.\*  
21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep.\*  
22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep.\*  
23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep.\*  
24: /cgn2\_6/ptodata/2/paa/US100\_COMB.pep.\*  
25: /cgn2\_6/ptodata/2/paa/US101\_COMB.pep.\*  
26: /cgn2\_6/ptodata/2/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1131	100.0	218	23	US-09-970-076-10
2	1131	100.0	297	18	US-09-488-725A-2121
3	966	85.4	274	18	PCT-US99-31025-50
4	966	85.4	274	18	US-09-471-179-50
5	966	85.4	301	1	PCT-US99-31025-132
6	966	85.4	301	18	US-09-471-179-132
7	966	85.4	306	1	PCT-US99-31025-30

ALIGNMENTS

RESULT 1

US-09-970-076-10  
; Sequence 10, Application US/09970076  
; GENERAL INFORMATION:  
; APPLICANT: Young, John A.T.  
; APPLICANT: Bradley, Kenneth A.  
; APPLICANT: Collier, Robert J.  
; APPLICANT: Mogridge, Jeremy S.  
; TITLE OF INVENTION: Anthrax Toxin Receptor  
; FILE REFERENCE: 960296.97745  
; CURRENT APPLICATION NUMBER: US/09/970, 076  
; PRIOR FILING DATE: 2001-10-03  
; PRIOR APPLICATION NUMBER: 60/251,481  
; PRIOR FILING DATE: 2000-12-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 218  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-970-076-10

Query Match 100.0%; Score 1131; DB 23; Length 218;  
Best Local Similarity 100.0%; Pred. No. 4.5e-112;  
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSFIVFSTRGTTLMKLTEDREQIRQGLEELQKVLPGDYMHEGFRASEQIYYENRQY 60  
|||||  
DB 1 MSFIVFSTRGTTLMKLTEDREQIRQGLEELQKVLPGDYMHEGFRASEQIYYENRQY 60

8 85.4 306 1 PCT-US99-31025-51 Sequence 51, Appl  
9 966 85.4 306 18 US-09-471-179-30 Sequence 30, Appl  
10 966 85.4 306 18 US-09-471-179-51 Sequence 51, Appl  
11 966 85.4 333 1 PCT-US99-31025-9 Sequence 9, Appl  
12 966 85.4 333 16 US-09-223-546-9 Sequence 9, Appl  
13 966 85.4 333 18 US-09-471-179-9 Sequence 9, Appl  
14 966 85.4 333 19 US-09-599-596-9 Sequence 9, Appl  
15 966 85.4 333 21 US-09-796-753-12 Sequence 12, Appl  
16 966 85.4 333 23 US-09-970-076-8 Sequence 8, Appl  
17 966 85.4 368 23 US-09-970-076-2 Sequence 2, Appl  
18 966 85.4 403 1 PCT-US00-30045-94 Sequence 94, Appl  
19 966 85.4 403 1 PCT-US01-11988-621 Sequence 621, Appl  
20 966 85.4 403 22 US-09-833-245-621 Sequence 621, Appl  
21 966 85.4 564 23 US-09-918-715-187 Sequence 187, Appl  
22 966 85.4 564 23 US-09-918-715-232 Sequence 232, Appl  
23 966 85.4 564 23 US-09-970-076-6 Sequence 6, Appl  
24 961 85.0 403 1 PCT-US00-30045-125 Sequence 125, Appl  
25 961 85.0 403 1 PCT-US01-11988-620 Sequence 620, Appl  
26 961 85.0 403 22 US-09-833-245-620 Sequence 620, Appl  
27 960 84.9 562 23 US-09-918-715-194 Sequence 194, Appl  
28 960 84.9 562 23 US-09-918-715-301 Sequence 301, Appl  
29 699 61.8 172 1 PCT-US99-31025-105 Sequence 105, Appl  
30 699 61.8 172 18 US-09-471-179-105 Sequence 105, Appl  
31 506 44.7 272 1 PCT-US01-16450-1574 Sequence 1574, Ap  
32 506 44.7 272 1 PCT-US01-16450A-1574 Sequence 1574, Ap  
33 506 44.7 284 1 PCT-US00-05226-27 Sequence 27, Appl  
34 506 44.7 284 19 US-09-516-745-27 Sequence 27, Appl  
35 506 44.7 455 1 PCT-US00-05226-26 Sequence 26, Appl  
36 506 44.7 455 19 US-09-516-745-26 Sequence 26, Appl  
37 506 44.7 487 1 PCT-US00-05226-107 Sequence 107, Appl  
38 506 44.7 487 19 US-09-516-745-107 Sequence 107, Appl  
39 506 44.7 488 1 PCT-US00-05226-2 Sequence 2, Appl  
40 506 44.7 488 1 PCT-US00-05226-57 Sequence 57, Appl  
41 506 44.7 488 1 PCT-US00-05226-59 Sequence 59, Appl  
42 506 44.7 488 1 PCT-US00-05226-61 Sequence 61, Appl  
43 506 44.7 488 19 US-09-516-745-2 Sequence 2, Appl  
44 506 44.7 488 19 US-09-516-745-57 Sequence 57, Appl  
45 506 44.7 488 19 US-09-516-745-59 Sequence 59, Appl

; APPLICANT: Millennium Pharmaceuticals, Inc.			
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS			
; FILE REFERENCE: 7853-173-228			
; CURRENT APPLICATION NUMBER: PCT/US99/31025			
; CURRENT FILING DATE: 1999-12-23			
; EARLIER APPLICATION NUMBER: 09/223,546			
; EARLIER FILING DATE: 1998-12-30			
; NUMBER OF SEQ ID NOS: 135			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 50			
; LENGTH: 274			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
; PCT-US99-31025-50			
Query Match		85.4%;	Score 966; DB 1; Length 274;
Best Local Similarity		98.9%;	Pred. No. 2.9e-94;
Matches 188;		Conservative 2;	Mismatches 0; Indels 0; Gaps 0;
QY	1	MSFVSTRGTTLMKLTEDREQIRQGLEELQKVLPGGDTYMHGEGFERASEQIYYENRQY	60
Db	53	MSFVSTRGTTLMKLTEDREQIRQGLEELQKVLPGGDTYMHGEGFERASEQIYYENRQY	112
QY	61	RTASVVIALTGDGELHEDLFFYSEREANRSDLGAIYVCVKDFNETQIARIADSKDHVF	120
Db	113	RTASVVIALTGDGELHEDLFFYSEREANRSDLGAIYVCVKDFNETQIARIADSKDHVF	172
QY	121	PVNDGFOALQGIHHSILKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSF	180
Db	173	PVNDGFOALQGIHHSILKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSF	232
QY	181	KINDSVTLKSLQSPWVSSTSGFEGNSHPCLPARPH	218
Db	233	KINDSVTLKSLQSPWVSSTSGFEGNSHPCLPARPH	242
RESULT 4			
US-09-471-179-50			
; Sequence 50, Application US/094711179			
; GENERAL INFORMATION:			
; APPLICANT: Holtzman, Douglas			
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM			
; FILE REFERENCE: 7853-173			
; CURRENT APPLICATION NUMBER: US/09/471,179			
; CURRENT FILING DATE: 1999-12-23			
; PRIOR APPLICATION NUMBER: 09/223,546			
; PRIOR FILING DATE: 1998-12-30			
; NUMBER OF SEQ ID NOS: 135			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 50			
; LENGTH: 274			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
; US-09-471-179-50			
Query Match		85.4%;	Score 966; DB 18; Length 274;
Best Local Similarity		98.9%;	Pred. No. 2.9e-94;
Matches 188;		Conservative 2;	Mismatches 0; Indels 0; Gaps 0;
QY	1	MSFVSTRGTTLMKLTEDREQIRQGLEELQKVLPGGDTYMHGEGFERASEQIYYENRQY	60
Db	53	MSFVSTRGTTLMKLTEDREQIRQGLEELQKVLPGGDTYMHGEGFERASEQIYYENRQY	112
QY	61	RTASVVIALTGDGELHEDLFFYSEREANRSDLGAIYVCVKDFNETQIARIADSKDHVF	120
Db	113	RTASVVIALTGDGELHEDLFFYSEREANRSDLGAIYVCVKDFNETQIARIADSKDHVF	172
QY	121	PVNDGFOALQGIHHSILKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSF	180
Db	173	PVNDGFOALQGIHHSILKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSF	232
Query Match		100.0%;	Score 1131; DB 18; Length 297;
Best Local Similarity		100.0%;	Pred. No. 7.1e-112;
Matches 218;		Conservative 0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	MSFVSTRGTTLMKLTEDREQIRQGLEELQKVLPGGDTYMHGEGFERASEQIYYENRQY	60
Db	80	MSFVSTRGTTLMKLTEDREQIRQGLEELQKVLPGGDTYMHGEGFERASEQIYYENRQY	139
QY	61	RTASVVIALTGDGELHEDLFFYSEREANRSDLGAIYVCVKDFNETQIARIADSKDHVF	120
Db	140	RTASVVIALTGDGELHEDLFFYSEREANRSDLGAIYVCVKDFNETQIARIADSKDHVF	199
QY	121	PVNDGFOALQGIHHSILKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSF	180
Db	200	PVNDGFOALQGIHHSILKSCIEILAAEPSTICAGESFQVVVRGNGFRHARNVDRVLCSF	259
QY	181	KINDSVTLKSLQSPWVSSTSGFEGNSHPCLPARPH	218
Db	260	KINDSVTLKSLQSPWVSSTSGFEGNSHPCLPARPH	297
RESULT 3			
PCT-US99-31025-50			
; Sequence 50, Application PC/TUS9931025			
; GENERAL INFORMATION:			

```
Qy 181 KINDSVTLK 190
Db 233 KINDSVTLNE 242

RESULT 5
PCT-US99-31025-132
; Sequence 132, Application PC/TUS9931025
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: 7853-173-228
; CURRENT APPLICATION NUMBER: PCT/US99/31025
; CURRENT FILING DATE: 1998-12-23
; EARLIER APPLICATION NUMBER: 09/223,546
; EARLIER FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 132
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US99-31025-132

Query Match 85.4%; Score 966; DB 1; Length 301;
Best Local Similarity 98.9%; Pred. No. 3.3e-94;
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSFVFSRGTTLMKLTEDREQIRGLEELQKLVLPGGDTYMHGFERASEQIYYENRQY 60
Db 80 MSFVFSRGTTLMKLTEDREQIRGLEELQKLVLPGGDTYMHGFERASEQIYYENRQY 139

Qy 61 RTASVIIALTGDELHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQIARIADSKDHVF 120
Db 140 RTASVIIALTGDELHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQIARIADSKDHVF 199

Qy 121 PVNDGFQALOGIIHSILKSKCIEILAAEPSTICAGESFQVVVVRGNGFRHARNVDRVLCSE 180
Db 200 PVNDGFQALOGIIHSILKSKCIEILAAEPSTICAGESFQVVVVRGNGFRHARNVDRVLCSE 259

Qy 181 KINDSVTLK 190
Db 260 KINDSVTLNE 269

RESULT 6
US-09-471-179-132
; Sequence 132, Application US/09471179
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 7853-173
; CURRENT APPLICATION NUMBER: US/09/471,179
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 132
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-471-179-132

Query Match 85.4%; Score 966; DB 18; Length 301;
Best Local Similarity 98.9%; Pred. No. 3.3e-94;
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSFVFSRGTTLMKLTEDREQIRGLEELQKLVLPGGDTYMHGFERASEQIYYENRQY 60
```

```
Db 80 MSFVFSRGTTLMKLTEDREQIRGLEELQKLVLPGGDTYMHGFERASEQIYYENRQY 139

Qy 61 RTASVIIALTGDELHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQIARIADSKDHVF 120
Db 140 RTASVIIALTGDELHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQIARIADSKDHVF 199

Qy 121 PVNDGFQALOGIIHSILKSKCIEILAAEPSTICAGESFQVVVVRGNGFRHARNVDRVLCSE 180
Db 200 PVNDGFQALOGIIHSILKSKCIEILAAEPSTICAGESFQVVVVRGNGFRHARNVDRVLCSE 259

Qy 181 KINDSVTLK 190
Db 260 KINDSVTLNE 269

RESULT 7
PCT-US99-31025-30
; Sequence 30, Application PC/TUS9931025
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: 7853-173-228
; CURRENT APPLICATION NUMBER: PCT/US99/31025
; CURRENT FILING DATE: 1999-12-23
; EARLIER APPLICATION NUMBER: 09/223,546
; EARLIER FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US99-31025-30

Query Match 85.4%; Score 966; DB 1; Length 306;
Best Local Similarity 98.9%; Pred. No. 3.4e-94;
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSFVFSRGTTLMKLTEDREQIRGLEELQKLVLPGGDTYMHGFERASEQIYYENRQY 60
Db 53 MSFVFSRGTTLMKLTEDREQIRGLEELQKLVLPGGDTYMHGFERASEQIYYENRQY 112

Qy 61 RTASVIIALTGDELHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQIARIADSKDHVF 120
Db 113 RTASVIIALTGDELHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQIARIADSKDHVF 172

Qy 121 PVNDGFQALOGIIHSILKSKCIEILAAEPSTICAGESFQVVVVRGNGFRHARNVDRVLCSE 180
Db 173 PVNDGFQALOGIIHSILKSKCIEILAAEPSTICAGESFQVVVVRGNGFRHARNVDRVLCSE 232

Qy 181 KINDSVTLK 190
Db 233 KINDSVTLNE 242

RESULT 8
PCT-US99-31025-51
; Sequence 51, Application PC/TUS9931025
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: 7853-173-228
; CURRENT APPLICATION NUMBER: PCT/US99/31025
; CURRENT FILING DATE: 1999-12-23
; EARLIER APPLICATION NUMBER: 09/223,546
; EARLIER FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
```

```
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US99-31025-51

Query Match      85.4%; Score 966; DB 1; Length 306;
Best Local Similarity 98.9%; Pred. No. 3.4e-94;
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFVSTRTGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFERASEQIYYENRQGY 60
Db 53 MSFVSTRTGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFERASEQIYYENRQGY 112
QY 61 RTASVITATDGHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQARIADSKDHVF 120
Db 113 RTASVITATDGHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQARIADSKDHVF 172
QY 121 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVDRVLCSE 180
Db 173 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVDRVLCSE 232
QY 181 KINDSVTLSK 190
Db 233 KINDSVTLINE 242

RESULT 9
PCT-US9-471-179-30
; Sequence 30, Application US/09471179
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 7853-173
; CURRENT APPLICATION NUMBER: US/09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-471-179-30

Query Match      85.4%; Score 966; DB 18; Length 306;
Best Local Similarity 98.9%; Pred. No. 3.4e-94;
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFVSTRTGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFERASEQIYYENRQGY 60
Db 53 MSFVSTRTGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFERASEQIYYENRQGY 112
QY 61 RTASVITATDGHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQARIADSKDHVF 120
Db 113 RTASVITATDGHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQARIADSKDHVF 172
QY 121 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVDRVLCSE 180
Db 173 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVDRVLCSE 232
QY 181 KINDSVTLSK 190
Db 233 KINDSVTLINE 242

RESULT 11
PCT-US99-31025-9
; Sequence 9, Application PC/TUS9931025
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS
; TITLE OF INVENTION: ENCODING THEM
; FILE REFERENCE: 7853-173-228
; CURRENT APPLICATION NUMBER: PCT/US99/31025
; CURRENT FILING DATE: 1999-12-23
; EARLIER APPLICATION NUMBER: 09/223,546
; EARLIER FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US99-31025-9

Query Match      85.4%; Score 966; DB 1; Length 333;
Best Local Similarity 98.9%; Pred. No. 3.8e-94;
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFVSTRTGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFERASEQIYYENRQGY 60
Db 80 MSFVSTRTGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFERASEQIYYENRQGY 139
QY 61 RTASVITATDGHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQARIADSKDHVF 120
Db 140 RTASVITATDGHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQARIADSKDHVF 199
QY 121 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVDRVLCSE 180
Db 200 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVDRVLCSE 259
```

```
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US99-31025-51

Query Match      85.4%; Score 966; DB 1; Length 306;
Best Local Similarity 98.9%; Pred. No. 3.4e-94;
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFVSTRTGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFERASEQIYYENRQGY 60
Db 53 MSFVSTRTGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFERASEQIYYENRQGY 112
QY 61 RTASVITATDGHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQARIADSKDHVF 120
Db 113 RTASVITATDGHEDLFFYSEREANRSDLGAIYVCVGVKDFNETQARIADSKDHVF 172
QY 121 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVDRVLCSE 180
Db 173 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFOVVVRGNGFRHARNVDRVLCSE 232
QY 181 KINDSVTLSK 190
Db 233 KINDSVTLINE 242

RESULT 10
US-09-471-179-51
; Sequence 51, Application US/09471179
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas
```

```
QY 181 KINDSVTLSK 190
Db 260 KINDSVTLINE 269

RESULT 12
US-09-223-546-9
; Sequence 9, Application US/09223546
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 09404/066001
; CURRENT APPLICATION NUMBER: US/09/223,546
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-223-546-9

Query Match      85.4%; Score 966; DB 16; Length 333;
Best Local Similarity 98.9%; Pred. No. 3.8e-94;
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFFRASEQIYYENRQGY 60
Db 80 MSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFFRASEQIYYENRQGY 139

QY 61 RTASVIALTDGELHEDLFFYSEREANRDLGAIVYCVGVKDFNETQLARIADSKDHVF 120
Db 140 RTASVIALTDGELHEDLFFYSEREANRDLGAIVYCVGVKDFNETQLARIADSKDHVF 199

QY 121 PVNDGFOALQGIHHSILKKSCIEILAAEPSTICAGSFQVYVVRGNGFRHARNVDRVLCSE 180
Db 200 PVNDGFOALQGIHHSILKKSCIEILAAEPSTICAGSFQVYVVRGNGFRHARNVDRVLCSE 259

QY 181 KINDSVTLSK 190
Db 260 KINDSVTLINE 269

RESULT 13
US-09-471-179-9
; Sequence 9, Application US/09471179
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 7853-173
; CURRENT APPLICATION NUMBER: US/09/471,179
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-471-179-9

Query Match      85.4%; Score 966; DB 18; Length 333;
Best Local Similarity 98.9%; Pred. No. 3.8e-94;
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFFRASEQIYYENRQGY 60
Db 80 MSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFFRASEQIYYENRQGY 139

QY 61 RTASVIALTDGELHEDLFFYSEREANRDLGAIVYCVGVKDFNETQLARIADSKDHVF 120
```

```
Db 140 RTASVIALTDGELHEDLFFYSEREANRDLGAIVYCVGVKDFNETQLARIADSKDHVF 199

QY 121 PVNDGFOALQGIHHSILKKSCIEILAAEPSTICAGSFQVYVVRGNGFRHARNVDRVLCSE 180
Db 200 PVNDGFOALQGIHHSILKKSCIEILAAEPSTICAGSFQVYVVRGNGFRHARNVDRVLCSE 259

QY 181 KINDSVTLSK 190
Db 260 KINDSVTLINE 269

RESULT 14
US-09-599-596-9
; Sequence 9, Application US/09599596
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas
; TITLE OF INVENTION: SECRETED PROTEINS AND NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 09404/066001
; CURRENT APPLICATION NUMBER: US/09/599,596
; CURRENT FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-599-596-9

Query Match      85.4%; Score 966; DB 19; Length 333;
Best Local Similarity 98.9%; Pred. No. 3.8e-94;
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFFRASEQIYYENRQGY 60
Db 80 MSFIVSTRGTTLMKLTEDREQIROGLEELQKVLPGDGYMHGFFRASEQIYYENRQGY 139

QY 61 RTASVIALTDGELHEDLFFYSEREANRDLGAIVYCVGVKDFNETQLARIADSKDHVF 120
Db 140 RTASVIALTDGELHEDLFFYSEREANRDLGAIVYCVGVKDFNETQLARIADSKDHVF 199

QY 121 PVNDGFOALQGIHHSILKKSCIEILAAEPSTICAGSFQVYVVRGNGFRHARNVDRVLCSE 180
Db 200 PVNDGFOALQGIHHSILKKSCIEILAAEPSTICAGSFQVYVVRGNGFRHARNVDRVLCSE 259

QY 181 KINDSVTLSK 190
Db 260 KINDSVTLINE 269

RESULT 15
US-09-796-753-12
; Sequence 12, Application US/09796753
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/122,458
```

; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/606,317
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 12
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-753-12

Query Match 85.4%; Score 966; DB 21; Length 333;
Best Local Similarity 98.9%; Pred. No. 3.8e-94;
Matches 188; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFIVSTRGTTLMKLTREIRQGLELEQKVLPGDYMHEGFERASEQIYYENRQGY 60
DB 80 MSFIVSTRGTTLMKLTREIRQGLELEQKVLPGDYMHEGFERASEQIYYENRQGY 139
QY 61 RTASVITATDGLHEDLFFYSERANRSRDLGAIIVYCVGKDFNETQLARIADSKDHVF 120
DB 140 RTASVITATDGLHEDLFFYSERANRSRDLGAIIVYCVGKDFNETQLARIADSKDHVF 199
QY 121 PVNDGFOALQGIITHSLKKSCIEILAAEPSTICAGESFQVVRGNGFRHARNVDRLVCSF 180
DB 200 PVNDGFOALQGIITHSLKKSCIEILAAEPSTICAGESFQVVRGNGFRHARNVDRLVCSF 259
QY 181 KINDSVTLTK 190
DB 260 KINDSVTLNE 269

Search completed: August 9, 2002, 10:44:32
Job time: 870 sec



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 8, 2002, 20:13:53 ; Search time 2803.4 Seconds

(without alignments)

6807.696 Million cell updates/sec

Title: US-09-970-076-1

Perfect score: 1414

Sequence: 1 aggccgcgcgaggagggcc.....aaaaaaaaaaaaaaaaaaaaa 1414

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_gss:\*
- 13: em\_gss\_hum:\*
- 14: em\_gss\_inv:\*
- 15: em\_gss\_pln:\*
- 16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	780.8	55.2	1063	9	AL542724
2	740.6	52.4	878	10	BI823853
3	688	48.7	963	10	BG326444
4	685.6	48.5	820	10	BG281561
5	681.4	48.2	1614	11	AK013005
6	637.8	45.1	725	10	BE741333
7	582	41.2	964	10	BG281831
8	493	34.9	513	9	BE146084
9	472.2	33.4	542	9	BE146075
10	459	32.5	487	9	BE146355
11	447.4	31.6	638	10	BE369415
12	432	30.6	569	10	BM050390
13	420.2	29.7	465	9	BE145958
14	390.2	27.6	693	10	BG323538
15	381.2	27.0	539	9	AI466645
16	373.6	26.4	554	9	AI084806
17	368	26.0	464	9	AU129171

18	329.8	23.3	356	9	BE146135
19	304.6	21.5	608	10	BI391348
20	295.4	20.9	369	9	AA004013
21	274	19.4	767	9	AI905553
22	266.6	18.9	527	9	AV667876
23	261.6	18.5	343	9	AI750523
24	247.8	17.5	282	9	BE146276
25	246.8	17.5	470	9	AA037321
26	242.4	17.1	842	10	BG576342
27	241.8	17.1	287	10	BG999787
28	239.8	17.0	360	10	BE478445
29	227.6	16.1	389	10	BF839348
30	224.8	15.9	441	10	BI183103
31	220.8	15.6	664	10	BM423073
32	208	14.7	380	9	AW207263
33	195	13.8	430	10	BE842906
34	192.8	13.6	340	10	BF370370
35	188.8	13.4	604	9	BE620906
36	188.4	13.3	245	10	BF370387
37	181	12.8	723	10	BI917586
38	175.6	12.4	672	10	BI869089
39	168.6	11.9	434	9	AV749755
40	167.8	11.9	341	9	BE146205
41	141.6	10.0	757	9	BB649806
42	137.2	9.7	562	9	AW106196
43	121.8	8.6	581	10	BM149748
44	121.2	8.6	301	10	BG951999
45	120.8	8.5	985	10	BF531402

## ALIGNMENTS

RESULT	1	AL542724	AL542724	1063 bp	mRNA	linear	EST 16-FEB-2001
AL542724	LOCUS	AL542724	LTI_FL002_PL1	Homo sapiens	cdna	clone	CS0DE011YL09 5 prime
DEFINITION							
ACCESSION		AL542724					
VERSION		AL542724.1	GI:12875049				
KEYWORDS		EST.					
SOURCE		human.					
ORGANISM		Homo sapiens					
REFERENCE		1 (bases 1 to 1063)					
AUTHORS		Li,W.B., Gruber,C., Jessee,J. and Polayes,D.					
TITLE		Full-length cdna libraries and normalization					
JOURNAL		Unpublished (2001)					
COMMENT		Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.					
FEATURES		Location/Qualifiers					
source		1..1063					
		/organism="Homo sapiens"					
		/db_xref="taxon:9606"					
		/clone="CS0DE011YL09"					
		/lab_host="DH10B"					
		/note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville , Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"					
BASE COUNT		272 a	254 c	314 g	221 t	2 others	
ORIGIN							

[illegible]

Qy 468 aaggattgaaagggccagtgagcagatttattatgaacacagacaaggtacagagac 527  
Db 421 AAGGATTTGAAGAGGCCAGTGAAGAGATTTATTATGAAACAGACAAGGTACAGACAG 480  
Qy 528 caagcgtcatcatattgcttctgactgtagagaaactccatgaagatctctttctattcaag 587  
Db 481 COAGCGTCAATTCCTTTGACTGATGAGAACTCCATGAAGATCTCTTTTCTATTTCAG 540  
Qy 588 agagggagcctaataaggctcagatcttggtgcaattgttactgtgtgtgtgaaag 647  
Db 541 AGAGGGAGGCTAATAGGCTCAGATCTGGTGCAATTTGTACTGTGT-TGTGTAAG 599  
Qy 648 atttcaatgagacacagctggcccgattggcgacagtaagatcatgtgtttcccgta 707  
Db 600 ATTTCAATGAGACACAGCTGGCCGGATTGCGACAGTAAGATCATGTGTTTCCCGTGA 659  
Qy 708 atgacggctttcagctctgcaagcaccatccactcaattttg-aagaagtcctgcatc 766  
Db 660 ATGACGGCTTTTCAGCTCTGCAAGGCATATCCACTCAATTTTGAAGAAGTCTTGCATC 719  
Qy 767 -gaaattctagcagctgaaccatccaccatattgtcagagagctc-atttcaagttgtcg 824  
Db 720 GGAATTTAGCAGCTGAACCATCACCATATGTCAGGAGAGTCAATTCGAAGTTGTCG 779  
Qy 825 tgagagg--aaacggcttcagacatgccg--caacgtggacaggg---tcctctgcagc 877  
Db 780 TCAGAGGGAACCGGTTTCCGACATGCGCGCAACACGTGGACCGGTCCTTGTGAGGT 839  
Qy 878 ttcagataaatgactcgtgcacatcaatgagaagccc 916  
Db 840 TTCCAGATCATGACTTCGGTCACACTTCAATGAAAAGCC 878

RESULT 3  
LOCUS BG326444 963 bp mRNA linear EST 27-FEB-2001  
DEFINITION 602425082F1 NIH\_MGC\_14 Homo sapiens cDNA clone IMAGE:4563020 5',  
mRNA sequence.  
ACCESSION BG326444  
VERSION BG326444.1 GI:13132881  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 963)  
NIH-MGC <http://mgi.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: DCTD/DNP  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCM1276 row: C column: 21  
High quality sequence stop: 718.  
Location/Qualifiers  
1..963  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4563020"  
/clone\_lib="NIH\_MGC\_14"  
/tissue\_type="renal cell adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: kidney; Vector: pOTB7; Site\_1: xhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGACAGAG(S). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in

FEATURES  
source

the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 261 a 217 c 292 g 193 t  
ORIGIN

Query Match 48.7%; Score 688; DB 10; Length 963;  
Best Local Similarity 96.8%; Pred. No. 9.6e-86;  
Matches 724; Conservative 0; Mismatches 20; Indels 4; Gaps 2;

Qy 1 aggaacccgcaggaagggcccgatggcgctccctgaggggtcgtgaggttcgaggttcgag 60  
Db 11 AGGACCCCGCAGGAAGGGCCCGGATGGCGCTCCCTGAGGTCGTGGCGAGTTCCGG 70  
Qy 61 aqcggtggaagagcggaccctctctcccggtcggccatgagccacgagcagagc 120  
Db 71 AGCGTGGGAAGGAGGCGACCCCTGCTCTCCCGGGCTGCGGGCCATGCGCCAGCGAGCG 130  
Qy 121 gagagccctcgccatcggtctccagtggtctctctttggccactctggtgctcatctgcgc 180  
Db 131 GAGAGCCCTCGCATCGGCTTCCAGTGGCTCTCTTTGGCCACTCTGCTGCTCATCTGCGC 190  
Qy 181 cgggcaagggggacgcagggagatgggggtccagcctgctacgagggatttgacctga 240  
Db 191 CGGGCAAGGGGACGACGAGGAGTGGGGTCCAGCTGCTACGGCGGATTTGACCTGTA 250  
Qy 241 ctctattttggacaatcaggaagtgcctgcacactggaatgaaatctattacttgt 300  
Db 251 CTTTATTTTGGACAAATCAGGAAGTGTGCTGCACCACTGGAATGAAATCTATTACTTTGT 310  
Qy 301 ggaacagttggctcacaattcatcagccacagttgagaatgtctttattgtttctc 360  
Db 311 GGAACAGTTGGCTCACAATTCATCAGCCACAGTTGAGAATGTCCTTTATTGTTTCTC 370  
Qy 361 caccgcaggaacaaccttaataactgcagagaagacagagaacaaatccgtcaagcct 420  
Db 371 CACCCGAGGAACAACCTTAATGAAACTGACAGAAGACAGAGAACAATAATCCGTCAGGSCCT 430  
Qy 421 agaagaactccagaagtctccagggagagacactacatgcagagatttgaaag 480  
Db 431 AGAAGAACTCCAGAAAGTTCTCCAGAGGAGACACTTACATGCATGAAGGATTCGAAG 490  
Qy 481 gggcagtgagcagattattatgaacacagacaaggttacagagccagcagctcatc 540  
Db 491 GCCCAGTGAGCAGATTTATTATGAAAACAGACAAGGTCACAGGACAGCCGCTCATCAT 550  
Qy 541 tgccttgactgtagagaactccatgaagatctcttttctatctcagagagggagcctaa 600  
Db 551 TGCTTTGACTGATGGAGAACTCCATGAAGATCTCTTTTCTATTTCAGAGAGGAGGCTAA 610  
Qy 601 taggtctcagatcttggtgcaattgtttactgtgtgtgtgaaagatttcaatgagac 660  
Db 611 TAGGTCCTGAGATCTTGGTGCAATTTTACTGTGTGGTGTGAAAAGATCTCATGAGAC 670  
Qy 661 acagctggcccgagattgcgagcagtaagatcatgtgtttcccgtaagagcgttcca 720  
Db 671 AAGTTTGGCC--GGATGCGGACAGTAAGGATCATGTGTATCCCGTGAATGACGG--TTTA 726  
Qy 721 ggcctcgaagcagcatcatccactcaatt 748  
Db 727 AGATCTGGAAGGCATCATCCATTTCATT 754

RESULT 4  
LOCUS BG281561 820 bp mRNA linear EST 21-FEB-2001  
DEFINITION 602402412F1 NIH\_MGC\_20 Homo sapiens cDNA clone IMAGE:4544691 5',  
mRNA sequence.  
ACCESSION BG281561  
VERSION BG281561.1 GI:13030486  
KEYWORDS EST.  
SOURCE human.





JOURNAL  
COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgraphs-r@mail.nih.gov  
Tissue Procurement: DCTD/DTP  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LLCW810 row: n column: 19  
High quality sequence stop: 711.

FEATURES  
source

1. .725  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3948114"  
/clone\_lib="NIH\_MGC\_9"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: ovary; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 206 a 165 c 176 g 178 t  
ORIGIN

Query Match 45.1%; Score 637.8; DB 10; Length 725;  
Best Local Similarity 98.8%; Pred. No. 8.5e-79;  
Matches 653; Conservative 0; Mismatches 7; Indels 1; Gaps 1;  
QY 395 gacagagaaacatccctcaaggcctagaagaactccagaaagtctgcaggaggagac 454  
Db 1 GACAGAGAAACAAATCCCTCAAGCCCTAGAGAACTCAGAAAGTCTGCCAGGAGGAGAC 60  
QY 455 acttcatgatgaagattgaaaggccagtgagcagatttattatgaaacagacaa 514  
Db 61 ACTTACATGATGATGAGGATTTGAAAGGCCAGTGAGCAGATTATTATGAAACAGACAA 120  
QY 515 gggtagcaggacagcgtcattcattgcttactgatgagaaactccatgaagatctc 574  
Db 121 GGGTACAGGACAGCCGCTCATCTGCTTACTTGATGGAGAACTCCATGAAGATCTC 180  
QY 575 tttttcttcagagaggaggttaattagttctcagatcttggtgcaattgttactgt 634  
Db 181 TTTTTCATTACAGAGAGGAGGCTTAATAGGTCTCGAGATCTTGGTGCATTTGTTTACTGT 240  
QY 635 gttggtgtaaaagatttcaatgagacacagctggcccgattgctggcagcagtaagatcat 694  
Db 241 GTTGGTGTGAAGATTTCATGAGACACAGCTGCCCGGATTCGGGACAGTAAGATCAT 300  
QY 695 gtgtttcccgatgaatgcggtcttccaggtctgtgaaggcatccactccactcaatttgaag 754  
Db 301 GTGTTTCCCGTGAATGACGGCTTTCAGGCTCTGCAAGGCATCATCCACTCAATTTTGAAG 360  
QY 755 aagtcctgcataaattctcagctgaaccatccaccataatgtgcaggagagtcattt 814  
Db 361 AAGTCCCTGCATCGAAATTCATGAGCTGAAACCATCCACCATATGTGCAAGAGAGTCATTT 420  
QY 815 caagtgtcgtgagagaaacggcttccgacatgcccgcaacgtggacagggctcctctgc 874  
Db 421 CAAGTTGCTGTGAGAGAAACGGCTTCCGACATGCCCGCAACGTGGACAGGGTCTCTGC 480  
QY 875 agcttcaagtcaatgactcgttcaactcaatgagagcccttttctgtggaagaacct 934  
Db 481 AGCTTCAAGATCAATGACTCGGTACACTCAATGAGAAGCCCTTTTCTGTGGAAGATACT 540  
QY 935 tattactgttcagcgcctactctaa-aagaagttggcatgaaagctcactccaggt 993

Db

541 TATTACTGTTCACGCGCCTATCTTAACAGAAGTTGGCATGAAGCTGCACATCCAGGT 600  
QY 994 cagcatgaacgagtcctctctttttatctcagttctgtcatcatcaccacacacactg 1053  
Db 601 CAGCATGAACGAGGCGCTCTCTTTTATCTCCAGTTCTGTATCATCATCACCACACACTTG 660  
QY 1054 t 1054  
Db 661 T 661

RESULT 7  
BG281831

LOCUS 602403057F1 NIH\_MGC\_20 Homo sapiens cDNA clone IMAGE:4545302 5',  
DEFINITION mRNA sequence.  
ACCESSION BG281831  
VERSION BG281831.1 GI:13030757  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 964)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgraphs-r@mail.nih.gov  
Tissue Procurement: ATCC/DCTD/DTP  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LLCM1230 row: a column: 15  
High quality sequence stop: 659.

FEATURES  
source

1. .964  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4545302"  
/clone\_lib="NIH\_MGC\_20"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 237 a 241 c 283 g 203 t  
ORIGIN

Query Match 41.2%; Score 582; DB 10; Length 964;  
Best Local Similarity 98.3%; Pred. No. 3.6e-71;  
Matches 641; Conservative 0; Mismatches 5; Indels 6; Gaps 5;

QY 1 agaacccgcgagaaagggcccgagatggcgctccctgagggctcgtggcgagttcgcg 60  
Db 22 AGGACCCGCGAGGAAGGCGCGGATGGCGCGTCCCTGAGGGTCTGTCGGAGTTCGCGG 81  
QY 61 agcgtggagagcagcagccctactctcccgggctcagggccatgcccac-ggcgagc 119  
Db 82 AGCGTGGAAAGAGCGAGCGACCTCTCTCCCCGGCTCGGGCCCAATGCCACTGCGGAGC 141  
QY 120 ggagagccctcggcatcggtctccagtggctctctttggccactctggtgctcatcgcg 179  
Db 142 GGAGAGCCCTCGGCATCGGCTTCAGTGGTCTCTTTGGCCACTCTGTTGCTCATCTCGC 201







```

QY 205 tgggggtccagcctgctacggcgagattgacctgttacttccatttttgacacaaatccaggaag 264
Db 299 TGGGGGTCCAGCCTGCTACGGCGGATTTCACCTGTACTTCATTTTCGACAAATCAGGAAG 240
QY 265 ttgtctgcaccactggaatgaatctattctattcttggtaacagttggtgcacaaattcat 324
Db 239 TGTGCTGTCACCACTGGAATGAATCTATTACTTTGTGGAAACAGTTGGCTCACAATATCAT 180
QY 325 cagccacagttgagaaatctcttattgtttttctccaccggaggaacaccttaataa 384
Db 179 CAGCCACAGTTGAGAAATGCTCTTATTGTTTCTCCACCGAGGAACAGCCTTAATGAA 120
QY 385 actgacagaagacagagaacaaatccgtcaagccttagaagaaactcagaaagtctgcc 444
Db 119 ACTGACAGAAGACAGAGAACAATCGTCAGGCGCTAGGAAGAACTCCAGAAAGTTTGCC 60
QY 445 aggagagacactcatcatcagtaaggaattgaaagggccagtgagcagatttattatg 503
Db 59 AGGAGGAGACACTTACATGCATGAGGATTGAAA-GGCCAGTGTATCAGATTATTATTATG 2

RESULT 11
BE369415
LOCUS 601220816f1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3589875 5',
DEFINITION mRNA sequence.
ACCESSION BE369415
VERSION BE369415.1 GI:9314778
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM8756 row: P column: 04
High quality sequence stop: 573.
Location/Qualifiers
1. .638
/organism="Mus musculus"
/strain="CZECH II (feral)"
/db_xref="taxon:10090"
/clone="IMAGE:3589875"
/clone_lib="NCI_CGAP_Lu29"
/tissue.type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/Note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 173 a 148 c 173 g 144 t
ORIGIN

Query Match 31.6%; Score 447.4; DB 10; Length 638;
Best Local Similarity 85.8%; Pred. No. 1.3e-52;
Matches 543; Conservative 0; Mismatches 86; Indels 4; Gaps 4;

QY 257 tcagggaagtgtctgcaccactggaatgaatctattcttggtaacagttgggtcac 316
|||||

```

```

Db 1 TCAGGAAGTGTGCTGCACCACCTGGAATGAAATCTACTACTTCTGGGAGCAGCTGGCTCAT 60
QY 317 aaattcatcagccacagittgaaatgtctttattgtttctccaccggaggaacaaacc 376
Db 61 AGATTTCATCAGCCCACTAGCTAAGGATGCTCTTCATTTGTCTTCTACTCGAGGCAACT 120
QY 377 Tlaataaaaactga-cagaagacagagaaacaaatccg-tcaagccttagaagaactccaga 434
Db 121 TTAATGAACTAAGCTGAGGACAGGGAACAGATCCGAGCAAGGCCCTAGAAGAGCTCCAGA 180
QY 435 aggtcttcagagagagacacttacatgaaagattgaaagggcagtgagcaga 494
Db 181 AAGTTCTGCAGGAGAGACACTTACATGACGAAGAGATTTGAGAGGGCCAGTGAGCAGA 240
QY 495 ttattatataaacacagaaggtacagacacagcgcgtcatcattgcttgaactgatg 554
Db 241 TTTACTATGAGAACATCAAGGATACAGGACAGCGAGCTCATCTCGGCTTGACCGATG 300
QY 555 gagaactccatgaagatctcttttctattcagagagggaggtcctaataggtctcagatc 614
Db 301 GGGAGCTGCACGAGGAGACCTCTTCTTCTACTCAGAGAGGAGGCTAAACAGATCCGAGACC 360
QY 615 ttggtgcaattgttactgtgtgtgtaaaagatttcaatgagacacagctggcccgga 674
Db 361 TTTGGTCCGA-TGTTTACTGCGTTGGCGTGAAGGATTTCAATGAAGACTCAGTTGGCTCGGA 419
QY 675 ttgcgacacgaagatcatgtgtttccgtgtaacagcgtttcaggtctcgaaggca 734
Db 420 TTGCACACAGTAAGGACACAGCTGTTCTCTGTGAACACAGCGCTTCCAGGCTCTCCAAGGCA 479
QY 735 tcatccactcaatttgaagaagtcctgcatcgaaattctagcagctgaaccatccacca 794
Db 480 TTTATCCACTCAATTTTAAAGAAATCTTGCATCGAAATCTTGGCGGCTGAACCATCCACCA 539
QY 795 tatgtcagagagagtcatttcaagttgtcgtgagaggaacacgcttcacacatgcgcgca 854
Db 540 TCTGCGCGGAGAGTCTCTTCAAGTGTGCTGAAGAGAAATGGCTTTCCGACATGTCCGC- 598
QY 855 acgtggacagggctcctcgcagcttcaagatca 887
Db 599 ATGTGGACAGGGCCCTCTGGGCTCTAAATATGA 631

RESULT 12
BM050390
LOCUS 603632635f1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:5423217 5',
DEFINITION mRNA sequence.
ACCESSION BM050390
VERSION BM050390.1 GI:16779657
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1882 row: e column: 10
High quality sequence stop: 569.
Location/Qualifiers
1. .569
/organism="Homo sapiens"
source

```

```

/db_xref="taxon:9606"
/clone="IMAGE:5423217"
/clone_lib="NIH_MGC_43"
/tissue_type="normal pigmented retinal epithelium"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library. I"
BASE COUNT 139 a 140 c 181 g 109 t
ORIGIN

```

```

Query Match      30.6%; Score 432; DB 10; Length 569;
Best Local Similarity 95.7%; Pred. No. 1.8e-50;
Matches 511; Conservative 0; Mismatches 10; Indels 13; Gaps 6;

```

```

Qy 1 aggaccgcgagaaagccgcggatggcgctccctgagggctggtggcgagttcgcg 60
Db 41 AGGACCCGCGAGAAAGGCGCGGATGGCGCTCCTTGAGGGTCTGGCGAGTTGCGGG 100
Qy 61 agcgtgggaagcggagaccctctctcccgggctgcggccatggccacggcgagcg 120
Db 101 AGCGTGGGAAGGAGCGGACCCCTCTCTCCCGGGCTCGCGCCATGCCACGCGGAGCG 160
Qy 121 gagagccctcggtcggctccagtggtctctttggccactctg-gtgcctcatctgg 179
Db 161 GAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTTTTGGCCACTCTGTGTCTATCTGGC 220
Qy 180 ccgggcaaggggacgcaggaggtatgggggtccagcctgctacgcggattgaacct 239
Db 221 CCGGGCAAGGGGACGACGAGGATGGGGTCCAGCCTGTACGGCGGATTGACCTGT 280
Qy 240 actctatttggacaatcagaaagtgtgtcacacactggaaatgaatctattcttg 299
Db 281 ACTTCATTTTGGACAA----GAAGTGTGTGTGCACACCTGGGAATGAAATCTATTACTTTG 335
Qy 300 tggacagtttggtccacaattc-atcagccacagattgagaatgtcccttattgttcc 358
Db 336 TGGAAACAGTTGGGTCAACAATCAATCAGCCACACAGTTTGAGAAATGTCCTTTATGTTTC 395
Qy 359 tcaccgcgaggaacaccttaataatgaactgacagaagacagagaacaatacgtcaagc 418
Db 396 TCACCCGAGGAAACACCTTAATGAACCTGACAGAAGACAGAGAACAATCGGTCAAGGC 455
Qy 419 ctagaagaactccagaagtcttccaggagga-gacacttacatgcattgaaggtattg- 476
Db 456 CTAGAAGAAGTCTCCAGAAAGTCTCCAGGAGGACGACACTTACATGCAATGAAGGATTGA 515
Qy 477 -aaagggccagtgagcaga---tttattatgaaaacagacagaggtacaggaca 526
Db 516 CACAGGGCCAGTGACGAGACTCTATTATGAACAACACAGCAAGGCTACAGGACA 569

```

```

RESULT 13
LOCUS BE145958 465 bp mRNA linear EST 21-JUN-2000
DEFINITION MRO-HT0209-100100-103-f09 HT0209 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE145958
VERSION BE145958.1 GI:8608682
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 465)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,

```

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-MRO-HT0209-100>)  
 100-103-f09st3=2000-01-10st4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 12  
 High quality sequence stop: 464.  
 Location/Qualifiers  
 1. 465

#### FEATURES

```

source
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone_lib="HT0209"
  /dev_stage="Adult"
  /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
  Site_2: SmaI; A mini-library was made by cloning products
  derived from ORESTES PCR (U.S. Letters Patent application
  No. 196,716 - Ludwig Institute for Cancer Research)
  profiles into the pUC 18 vector. Reverse transcription of
  tissue mRNA and cDNA amplification were performed under
  low stringency conditions."
  low stringency conditions."
BASE COUNT 100 a 125 c 119 g 121 t
ORIGIN

```

```

Query Match      29.7%; Score 420.2; DB 9; Length 465;
Best Local Similarity 96.9%; Pred. No. 8.3e-49;
Matches 439; Conservative 0; Mismatches 13; Indels 1; Gaps 1;
Qy 71 ggagcgaccctgctct-ccccgggctgcggccatggccacgcgcggagagccct 129
Db 465 GGACGGACCCCTGTCTCCCCGGGCTCGGGCCATGGCCACCGCGAGAGGCCCT 406
Qy 130 cggcatcggtccagtggtctctttggccactctggtgctcatctgcgcgggcaag 189
Db 405 CGGCATCGGTTTCCAGTGGCTCTTTGGGCCACTCTGGTGTCTATCTGCCCGGCAAG 346
Qy 190 gggacgaggagatggggtccagcctgctacgcggatttgacctgtacttcat 249
Db 345 GGTACGCGAGGAGATGGGGTCTCAGCGCTGCTAGCGCGGATTGACCTGTACTTCATTT 286
Qy 250 ggacaaatcaggaagtgtgtgcaccactggaatgaaatctattcttggaaagtt 309
Db 285 GGACAAATCAGGAAGTGTGTGCACCACTGGAATGAAATCTATTACTTTGTGGAACAGTT 226
Qy 310 ggctcacaattcatcagccacagttgagaatgtcctttattgttttccaccag 369
Db 225 GGCTCAAAATTCATCAGCCACAGTTGAGAATGTCTCTTTATTGTTTTCCTCCACCCGAGG 166
Qy 370 acaacaccttaataaactgacagaagacagagaacaaatccgtcaaggccctagaagaact 429
Db 165 AACAACTTAATGAACCTACAGAAGACAGAGAACAATCCGTCAGGCCCTAGAGAAGT 106
Qy 430 ccagaagttctccaggaagacacacttacatgcattgaaggtattgaaagggccagtg 489
Db 105 CCAGAAAGTTCTCCAGGAGGACACTTACATGCATGAAGGATTTGAAAGGCCCAAGTGA 46
Qy 490 gcagattattatgaaacagacaaggggtacag 522

```

```
Db 45 GCAGATTATTATGAAAGTACAGAGGTACAG 13

RESULT 14
BG323538
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

602422195F1 NTH_MGC_14 Homo sapiens cDNA clone IMAGE:4560214 5',
mRNA sequence.
BG323538
BG323538.1 GI:13129975
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 693)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC1268 row: n column: 23
High quality sequence stop: 571.
Location/Qualifiers
1..693
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4560214"
/tissue_type="renal cell adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GCCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```
171 a 162 C 214 g 146 t
BASE COUNT
ORIGIN

Query Match 27.6%; Score 390.2; DB 10; Length 693;
Best Local Similarity 85.2%; Pred. No. 9.5e-45;
Matches 575; Conservative 0; Mismatches 63; Indels 37; Gaps 11;

QY 1 aggaccgcgaggaagggccc--gcggatggcgctccctgaggtcgtggcgagttcgc 58
|||||
Db 19 AGACCCCGCAGGAGGCGCCCTCGGTATGGCGCGCTCCCTGAGGCGTGGTGGCGAGTTCG 78
|||||
QY 59 ggaagctggaagagcggac-----cctgtctctcccggtcgcggccatggccacg 112
|||||
Db 79 GTGAGCTGTGGAGAGAGAGCTGAGCGCTGCTCTCTCCCGCGTGGCGCGTGGCCAT 138
|||||
QY 113 gcggagcg-----gagagccctcgagctcgcttccagtgactctctt 155
|||||
Db 139 TGTGACCGAGCGCAGGATCGGATGAGCGGCTCGGACATCGGCTTCACGTGGCTCTCT 198
|||||
QY 156 tggccacttggtgctcatctgcgcggggaagggagcagggaggtgggggtccag 215
|||||
Db 199 TGCCCACTCTGTGTCTCATCTGCGCCGCGCAAGGGGAGCAGGAGGAGTGGGGTCCAG 258
|||||
QY 216 cctgtacggcgaggttgcactgtacttctatttgacaaaatcagaagtgtctgcacc 275
|||||
Db 259 CTTGCTACGGCGGATTTGGACCTGTACTTCTATTTTGGGACAAATCAGGAAGTGTCTGCACC 318
|||||
```

```
276 actggaatgaatctattactttgtgaaacagttggctcacaattcatcagccacagt 335
|||||
319 ACTGGAATGAATCTATTACTTGGTGAACAGATTGGCTCACAATTCATCAGCCACAGT 378
|||||
336 tgaagtgtctcttattgttttccaccgcgaggaacaaccttaataactgacagaag 395
|||||
379 TGAGAATGTCTTATGTGTTTCTCCACCCGAGGAAACACTTAATGAAGAACTGACAGAAG 438
|||||
396 acagagaacaataccgtcaaggcctagaagaacccagaa-aggcttcgcagga-ggaga 453
|||||
439 ACAGAGAACAATCCGTCGAAGGCTAGAGAAGCTCCAGAACAGTTCGCCAGGATGGAGA 498
|||||
454 cacttaccatgcataa-ggatttgaaggccagtgagcagatttatttgaataaacagac 512
|||||
499 CACTTACATGCATGAACGGATCTGAAAGGCCAGTGAGCAGAGATTATTATGAACACAGAC 558
|||||
513 aagggtacaggaagcagcagc--gtcatcattgtttgactgagtg----gagaactccatg 566
|||||
559 AAGGTACAGGACAGCAGCCGCTCCACCATTTGGCTTTGGACTTGATGCGGAGAACTCCATG 618
|||||
567 agatctcttttctattca-gagaggaggagcctaagt-aggctcgcagatctt-ggtgcaa 623
|||||
619 AAGATCTCTTAACTATTATCAGAGAGCGAGGCTAAATAAGGCTCGGAAATCTTGGGTGCAC 678
|||||
624 ttgttactgtgttg 638
|-|-|-|-|-|
679 TGCTTAACTGGGTG 693
|-|-|-|-|-|

RESULT 15
AI466645
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AI466645 539 bp mRNA linear EST 15-MAR-2000
mg81d04.y1 Soares mouse embryo NbMEL3.5 14.5 Mus musculus CDNA
clone IMAGE:439399 5', mRNA sequence.
AI466645
AI466645.1 GI:4315264
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 539)
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person
, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter
, E., Kohn, S., Shin, T., Jackson, F., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
This read is a RESEQUENCE of a previously sequenced mouse clone.
This read has been verified (found to hit its original self in the
correct orientation)
Putative full length read
vector to vector length is 709
MGI:264735
Seq primer: -40RP from Gibco
High quality sequence stop: 454
POLYA=No.
Location/Qualifiers
1..539
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:439399"
```

```

/clone_lib="Soares mouse embryo NDME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Patima Bonaldo. "
```

BASE COUNT					153 a					119 c					138 g					129 t				
ORIGIN																								
Query Match					27.0%;					Score 381.2;					DB 9;					Length 539;				
Best Local Similarity					86.9%;					Pred. No. 1.8e-43;														
Matches 419;					Conservative 0;					Mismatches 63;					Indels 0;					Gaps 0;				
QY	227	ggatttgacctgtacttcatttggacaacatcaggaaagtgtgtgcaccactggaatgaa	286																					
Db	1	GGATTCGACCTCTACTTCCTCGGACAGTCAGGAAGTGTGTGCACCACTGGAATGAA	60																					
QY	287	atctattactttgtggaacagttggctcacaatttcattcatccacagttgagaatgtcc	346																					
Db	'61	ATCTACTACTTCGFGAGCAGTTGGCTCATAGATTATCATGCCCCACAGCTAAGGATGTCC	120																					
QY	347	tttattgttttccaccocaggagaacaccttaataaactgcagagaacagagaacaa	406																					
Db	121	TTCAATGCTCTCTCTACTCGAGGAGCAACTTTAATGAACTAACTGAGGACAGGGAACAG	180																					
QY	407	atccgtcaagcctagaagaactccagaagtctgtccaggaggagagacattacatgat	466																					
Db	181	ATCCGACAGGCCCTAGAAGAGCTCCAGAAAGTTCTGCCAGGAGGAGACACTTACATGCAC	240																					
QY	467	gaaggatttgaaaggccagtgagcagatttattatgaaacacagacaagggtacaggaca	526																					
Db	241	GAAGGATTTCGAGAGGCCAGTGAGCAGATTACTATGAGAACAGTCAAGGATACAGGACG	300																					
QY	527	gccagcgtcatcattgttactgtatggagaactccatccatgagagatctctttttattca	586																					
Db	301	GCGAGCGTCATATCGCGTTGAGGATGGGAGCTGCACGAGGACCTCTTCTTACTCA	360																					
QY	587	gagaggaggctaatagggtctcgagatcttgggtgcaattgtttactgtgtgtgtgaa	646																					
Db	361	GAGAGGGAGGCTAACCGATCCCGAGACCTTGGTGGGATTTGTTACTGCTGGCGTGAAG	420																					
QY	647	gatttcaatgagacacagctgcccggattgcgagcagtaaggatcgtgtttcccggtg	706																					
Db	421	GATTTCAATGAACACTCAGTTGGCTCGGATTCGAGACAGTAAAGGACCACTGTTCCTGTG	480																					
QY	707	aa 708																						
Db	481	AA 482																						

Search completed: August 8, 2002, 23:02:53  
Job time: 10140 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: August 8, 2002, 22:16:28 ; Search time 413.49 Seconds  
(without alignments)  
5871.284 Million cell updates/sec

Title: US-09-970-076-1  
Perfect score: 1414  
Sequence: 1 aggaccgcgaggaaggcc.....aaaaaaaaaaaaaaaaaaaaa 1414

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 347872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1193.6	84.4	2447	22 AAD05303	Human secreted pro
2	1192.4	84.3	2086	22 AAD05334	Human secreted pro
3	1052.4	74.4	2272	21 AAA47455	Human TANGO 197 co
4	906.8	64.1	1402	22 AA159918	Human polynucleoti
5	906.8	64.1	1609	22 AA158132	Human polynucleoti
6	890	62.9	1436	22 AAH14331	Human cDNA sequenc
7	453.2	32.1	4417	21 AAA47479	Murine TANGO 197 c
8	368	26.0	464	22 AAH06188	Human cDNA clone (
9	349	24.7	356	22 AAL19043	Human breast cance

10	348	24.6	355	22 AAL08133	Human breast cance
11	346	24.5	346	22 AAL25878	Human breast cance
12	345.8	24.5	3677	21 AAA75149	cDNA encoding a hu
13	344.2	24.3	3677	21 AAA75157	CDNA clone encodin
14	344.2	24.3	3677	21 AAA75158	CDNA clone encodin
15	344.2	24.3	3677	21 AAA75159	CDNA clone encodin
16	341.8	24.2	3501	21 AAA75162	CDNA clone encodin
17	341.2	24.1	2239	22 AAS31233	Human cDNA encodin
18	340.2	24.1	3501	21 AAA75150	cDNA encoding a mu
19	340.2	24.1	3501	21 AAA75160	CDNA clone encodin
20	338.6	23.9	3501	21 AAA75161	CDNA clone encodin
21	233.6	16.5	350	21 AAZ94448	Cartilage-associat
22	186.4	13.2	211	20 AAH86628	EST clone CG874.
23	113	8.0	811	22 AAH08055	Human cDNA clone (
24	113	8.0	1189	22 AAH15137	Human immune/haema
25	96.8	6.8	390	22 AAH57088	Human immune sequenc
26	59.6	4.2	188	21 AAA45398	Human secreted exp
27	59.4	4.2	6070	24 ABL33679	Human immune syste
28	59.4	4.2	6070	24 ABL34579	Human metastasis a
29	59.4	4.2	6668	24 ABL33696	Human immune syste
30	59.4	4.2	13712	24 ABL33531	Human immune syste
31	59.2	4.2	9219	22 AAS46807	Tumour suppressor
32	59.2	4.2	9219	24 ABL34426	Human immune syste
33	58.2	4.1	612	22 AAH71471	Human cervical can
34	58	4.1	6283	24 ABL32835	Human immune syste
35	57.6	4.1	7076	24 ABL34145	Human immune syste
36	57.6	4.1	11015	24 AAD22335	Chemically treated
37	56.8	4.0	5845	24 ABL33662	Human immune syste
38	56.2	4.0	5134	22 AAS45459	Chemically pretrea
39	56.2	4.0	5134	24 ABL33818	Human immune syste
40	56.2	4.0	7046	24 ABL34116	Human immune syste
41	56.2	4.0	8952	22 AAS46446	Tumour suppressor
42	56.2	4.0	12393	24 ABL33262	Human immune syste
43	56.2	4.0	18011	24 ABL32034	Human immune syste
44	56	4.0	1937	21 AAA47445	Maceque TANGO 232
45	55.8	3.9	6485	22 AAS46560	Tumour suppressor

ALIGNMENTS

RESULT 1

AAD05303

ID AAD05303 standard; cDNA: 2447 BP.

XX AC AAD05303;

XX DT 17-JUL-2001 (first entry)

XX DE Human-secreted protein-encoding gene 4

XX KW Human: secreted protein; proliferative disorder; cancer; tumour;  
foetal abnormality; developmental abnormality; haematopoietic disorder;  
immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
inflammation; allergy; neurological disorder; Alzheimer's disease;  
Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
cardiovascular disorder; angioneurotic disorder; kidney disorder;  
gastrointestinal disorder; pregnancy-related disorder;  
endocrine disorder; infection; wound healing; vulnerability;  
cell culture; chemotaxis; food additive; gene therapy;  
binding partner identification; chromosome 19; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 140..1351

XX FT /product= "Human secreted protein"

XX FT /tag= a

XX FT /tag= b

XX FT /tag= c

FT XX /product= "Mature human secreted protein"

PN XX WO200134626-A1.

PD XX 17-MAY-2001.

PP XX 01-NOV-2000; 2000WO-US30045.

XX XX 05-NOV-1999; 99US-0163581.

PR XX 30-JUN-2000; 2000US-0215133.

XX XX (HUMA-) HUMAN GENOME SCI INC.

PA XX

PI Ruben SM, Komatsoulis GA, Moore PA, Birse CE, Ni J;

XX P-PSDB; AAE01439.

DR WPI; 2001-308778/32.

XX P-PSDB; AAE01439.

XX New nucleic acid molecules encoding 28 human secreted proteins for

PT diagnosing, preventing, treating or ameliorating medical conditions and

PT used as food additives or preservatives -

XX

PS Claim 1; Page 425-426; 562pp; English.

XX AAD05300-AAD05379 represent cDNAs corresponding to 28 human secreted

CC protein genes, and AAE01436-AAE01513 represent the proteins they encode.

CC AAE01514-AAE01544 represent human secreted protein fragments or variants.

CC The genes and their secreted proteins are useful for preventing,

CC treating or ameliorating medical conditions, e.g., by protein or gene

CC therapy. Pathological conditions can be diagnosed by determining the

CC amount of the new protein in a sample or by determining the presence of

CC mutations in the new genes. Specific uses are described for each of the

CC 28 genes, based on the tissues in which they are most highly expressed,

CC and include developing products for the diagnosis or treatment of

CC proliferative disorders, cancer, tumours, foetal and developmental

CC abnormalities, haematopoietic disorders, diseases of the immune system,

CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,

CC allergies, neurological disorders (e.g., Alzheimer's disease,

CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,

CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,

CC cardiovascular disorders, angiogenic disorders, kidney disorders,

CC gastrointestinal disorders, pregnancy-related disorders, endocrine

CC disorders, and infections. The proteins can also be used to aid wound

CC healing and epithelial cell proliferation, to prevent skin aging due to

CC subburn, to maintain organs before transplantation, for supporting cell

CC culture of primary tissues, to regenerate tissues, to identify their

CC cognate ligands or binding partners, and in chemotaxis, and can be used

CC as a food additive or preservative to modify storage properties.

CC Antibodies specific for a protein of the invention can be used in

CC alleviating symptoms associated with the disorders mentioned above, and

CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked

CC immunosorbent assay (ELISA). The present sequence represents a human

CC secreted protein-encoding cDNA of the invention.

XX SQ Sequence 2447 BP; 673 A; 562 C; 587 G; 625 T; 0 other;

QY	181	cgggcaaggaggagcagcagagagatgggggtccagcctgctacgaggattgacctgta	240
Db	217	cgggcaaggaggagcagcagagagatgggggtccagcctgctacgaggattgacctgta	276
QY	241	cttcatttggacaaatcaggaagtgtgctgcaccactggaatgaaatctattactttgt	300
Db	277	cttcatttggacaaatcaggaagtgtgctgcaccactggaatgaaatctattactttgt	336
QY	301	ggaacagtgtggtcacaataattcattcagccacagttgagaatgtcctttattttcttc	360
Db	337	ggaacagtgtggtcacaataattcattcagccacagttgagaatgtcctttattttcttc	396
QY	361	cacccgagggaacaaccttaataatgaactgcagaagacagagaacaataccgtcaaggcct	420
Db	397	cacccgagggaacaaccttaataatgaactgcagaagacagagaacaataccgtcaaggcct	456
QY	421	agaagaactccagaagtctgcaggaggagacacttacatgcataaggaattgaaag	480
Db	457	agaagaactccagaagtctgcaggaggagacacttacatgcataaggaattgaaag	516
QY	481	ggccagtgcagcagatttattgaaacacagacagaggtacagacagccagcgtcatcat	540
Db	517	ggccagtgcagcagatttattgaaacacagacagaggtacagacagccagcgtcatcat	576
QY	541	tgcttgactgatggagaactccatgaagatctcttttttatttcagagaggaggctaa	600
Db	577	tgcttgactgatggagaactccatgaagatctcttttttatttcagagaggaggctaa	636
QY	601	taggtctcgagatctgtgtcaattgtttactgtgtgtgtgaaagatttcaatgagac	660
Db	637	taggtctcgagatctgtgtcaattgtttactgtgtgtgtgaaagatttcaatgagac	696
QY	661	acagctggcccgattgcggacagtaaggtatgtttcccggtgaatgacggcttca	720
Db	697	acagctggcccgattgcggacagtaaggtatgtttcccggtgaatgacggcttca	756
QY	721	ggctctgcaaggagatcatcacatcaatttgaagaagtcctgcatacgaattctagcgc	780
Db	757	ggctctgcaaggagatcatcacatcaatttgaagaagtcctgcatacgaattctagcgc	816
QY	781	tgaaccatccaccatattgcaggagagtcatttcaagtgtcgtgagagaaacgctt	840
Db	817	tgaaccatccaccatattgcaggagagtcatttcaagtgtcgtgagagaaacgctt	876
QY	841	ccgacatcccgcaacgtggacaggggtcctctgcagcttcaagaatcaatgactcgtc	900
Db	877	ccgacatcccgcaacgtggacaggggtcctctgcagcttcaagaatcaatgactcgtc	936
QY	901	actcaatgagaagccctttctgtggaagacacttatttactgtccagcgcctatctt	960
Db	937	actcaatgagaagccctttctgtggaagacacttatttactgtccagcgcctatctt	996
QY	961	aaaagaagtggcatgaaagctgcactccaggtcagcatgaacgatggcctcttttat	1020
Db	997	aaaagaagtggcatgaaagctgcactccaggtcagcatgaacgatggcctcttttat	1056
QY	1021	ctccagttctgtcatcatcacccacacacactgttctgacgggttccatctcctgcccacgc	1080
Db	1057	ctccagttctgtcatcatcacccacacacactgttctgacgggttccatctcctgcccacgc	1116
QY	1081	cctgtcatctgtctcctcctagccctggtcctcctcctcctcctcctcctcctcctc	1140
Db	1117	cctgtcatctgtctcctcctagccctggtcctcctcctcctcctcctcctcctcctc	1176
QY	1141	ctgcaactgtattatcaagagaggtccctccaccctcccgaggagagtgaggaaaaata	1200
Db	1177	ctgcaactgtattatcaagagaggtccctccaccctcccgaggagagtgaggaaaaata	1236

RESULT 2  
AAD05334  
ID AAD05334 standard; cDNA; 2086 BP.  
XX

AC AAD05334;  
 XX 17-JUL-2001 (first entry)  
 XX Human secreted protein-encoding gene 4 cDNA clone HWLFR02, SEQ ID NO:45.  
 XX Human; secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder;  
 KW endocrine disorder; infection; wound healing; vulnerability;  
 KW cell culture; chemotaxis; food additive; gene therapy;  
 KW binding partner identification; chromosome 19; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 166..1377  
 FT /\*tag= a  
 FT /product= "Human secreted protein"  
 FT /transl\_except= (pos:688..690, aa:Xaa)  
 FT /transl\_except= (pos:1123..1125, aa:Xaa)  
 FT /transl\_except= (pos:1156..1158, aa:Xaa)  
 FT /transl\_except= (pos:1267..1269, aa:Xaa)  
 FT /note= "Xaa equals any of the twenty naturally occurring  
 L-amino acids"  
 FT sig\_peptide 166..246  
 FT /\*tag= b  
 FT mat\_peptide 247..1374  
 FT /\*tag= c  
 FT /product= "Mature human secreted protein"  
 FT  
 XX  
 PN WO200134626-A1.  
 PD 17-MAY-2001.  
 XX  
 PF 01-NOV-2000; 2000WO-US30045.  
 XX  
 PR 05-NOV-1999; 99US-0163581.  
 PR 30-JUN-2000; 2000US-0215133.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ruben SM, Komatsoulis GA, Moore PA, Birse CE, Ni J;  
 XX  
 DR WPI: 2001-308778/32.  
 XX P-PSDB: AAE01469.  
 XX  
 PT New nucleic acid molecules encoding 28 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives -  
 XX  
 PS Claim 1: Page 449; 562pp; English.  
 XX  
 CC AAD05300-AAD05379 represent cDNAs corresponding to 28 human secreted  
 CC protein genes, and AAE01436-AAE01513 represent the proteins they encode.  
 CC AAE01514-AAE01544 represent human secreted protein fragments or variants.  
 CC The genes and their secreted proteins are useful for preventing,  
 CC treating or ameliorating medical conditions, e.g., by protein or gene  
 CC therapy. Pathological conditions can be diagnosed by determining the  
 CC amount of the new protein in a sample or by determining the presence of  
 CC mutations in the new genes. Specific uses are described for each of the  
 CC 28 genes, based on the tissues in which they are most highly expressed,  
 CC and include developing products for the diagnosis or treatment of  
 CC proliferative disorders, cancer, tumours, foetal and developmental  
 CC abnormalities, haematopoietic disorders, diseases of the immune system,  
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
 CC allergies, neurological disorders (e.g., Alzheimer's disease,  
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,

CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,  
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
 CC disorders, and infections. The proteins can also be used to aid wound  
 CC healing and epithelial cell proliferation, to prevent skin aging due to  
 CC sunburn, to maintain organs before transplantation, for supporting cell  
 CC culture of primary tissues, to regenerate tissues, to identify their  
 CC cognate ligands or binding partners, and in chemotaxis, and can be used  
 CC as a food additive or preservative to modify storage properties.  
 CC Antibodies specific for a protein of the invention can be used in  
 CC alleviating symptoms associated with the disorders mentioned above, and  
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
 CC immunosorbent assay (ELISA). The present sequence represents a human  
 XX secreted protein-encoding cDNA of the invention.  
 SQ Sequence 2086 BP; 544 A; 491 C; 538 G; 501 T; 12 other;  
 Query Match 84.3%; Score 1192.4; DB 22; Length 2086;  
 Best Local Similarity 99.4%; Pred. No. 5,5e-283;  
 Matches 1193; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 aggacccgcgaggaagggcccgcgatgcgctccctgagggctgctgaggttcgctgcgaggttcgcg 60  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 63 aggacccgcgaggaagggcccgcgatgcgctccctgagggctgctgaggttcgctgcgaggttcgcg 122  
 QY 61 acgtggtggaagagcgagccctgctctcccggttcgctgcgaggttcgctgcgaggttcgcg 120  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 123 agcgtggtggaagagcgagccctgctctcccggttcgctgcgaggttcgctgcgaggttcgcg 182  
 QY 121 gagagccctcgcatcgcttcacgtggctctctcttctggccactctggtgctcatctgcgc 180  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 183 gagagccctcgcatcgcttcacgtggctctcttctggccactctggtgctcatctgcgc 242  
 QY 181 cggggcaagggggacgcagggaggtgggggtccagcctgctacgctgcgaggttcgacctgta 240  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 243 cggggcaagggggacgcagggaggtgggggtccagcctgctacgctgcgaggttcgacctgta 302  
 QY 241 cttcatttttgacaaatcaggaagtgtgctgcaccactggaatgaatattattactttgt 300  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 303 cttcatttttgacaaatcaggaagtgtgctgcaccactggaatgaatattattactttgt 362  
 QY 301 ggaacagtttgctcacaatttcacagccacagtggttgagaatgtcttattgtttctc 360  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 363 ggaacagtttgctcacaatttcacagccacagtggttgagaatgtcttattgtttctc 422  
 QY 361 caccgaggaacaaccttaataatgaactgcagaagacagagacaataatcgtcaaggcct 420  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 423 caccgaggaacaaccttaataatgaactgcagaagacagagacaataatcgtcaaggcct 482  
 QY 421 agaagaactccagaagtctccagggagagacacacttacatgcataagagatttgaag 480  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 483 agaagaactccagaagtctccagggagagacacacttacatgcataagagatttgaag 542  
 QY 481 ggccagtgcagcattattattgaaacagacaaggttacaggacagcagcagcgtcatcat 540  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 543 ggccagtgcagcattattattgaaacagacaaggttacaggacagcagcagcgtcatcat 602  
 QY 541 tgccttgactgtaggagaactccatgaagatctcttttctatttcagagagggaggtctaa 600  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 603 tgccttgactgtaggagaactccatgaagatctcttttctatttcagagagggaggtctaa 662  
 QY 601 taggtctcagatcttggtgcattgtttactgtgtgtgtgaaagatttcaatagac 660  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 663 taggtctcagatcttggtgcattgtttactgtgtgtgtgaaagatttcaatagac 722  
 QY 661 acagctggccggttcgagcagtagtaaggatcgtgtttcccgtaaatgacggtcttca 720  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 723 acagctggccggttcgagcagtagtaaggatcgtgtttcccgtaaatgacggtcttca 782  
 QY 721 ggctctgacaggtcatcatccactcaatttttgaagaagtcctgcataatcttagcagc 780  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 783 ggctctgacaggtcatcatccactcaatttttgaagaagtcctgcataatcttagcagc 842

Qy	781	tgaaccatccaccatattgtcaggagagtgatcatttcaagtgtgtcgtgagagaaacgctt	840
Db	843	tgaaccatccaccatattgtcaggagagtgatcatttcaagtgtgtcgtgagagaaacgctt	902
Qy	841	ccgacatgccgcaacgtggacagggctcctcgtcagcttcaagatcaatgactcggtcac	900
Db	903	ccgacatgccgcaacgtggacagggctcctcgtcagcttcaagatcaatgactcggtcac	962
Qy	901	actcaatgagaagccctttctgtggaagacactatttactgtgtccagcgcctattt	960
Db	963	actcaatgagaagccctttctgtggaagacactatttactgtgtccagcgcctattt	1022
Qy	961	aaaagaagtgtgcatgaaagctgcactccaggtcagcatgacatgacgtcctcttttat	1020
Db	1023	aaaagaagtgtgcatgaaagctgcactccaggtcagcatgacatgacgtcctcttttat	1082
Qy	1021	ctcaagttctcattcatcacacacacactgttctgaagttccactcctggccatgc	1080
Db	1083	ctccagttctcattcatcacacacacactgttctgaagttccactcctggccatgc	1142
Qy	1081	cctgctgacatcctgtctcctagccctggtctctctctctgtgtgttctggccctctg	1140
Db	1143	cctgctgacatcctgtctcctagccctggtctctctctgtgtgttctggccctctg	1202
Qy	1141	ctgcaactgtattatcaaggaggtccctccaccctccagcagagagtgagaaataa	1200
Db	1203	ctgcaactgtattatcaaggaggtccctccaccctccagcagagagtgagaaataa	1262
RESULT 3			
AA47455	ID AA47455 standard; cDNA; 2272 BP.		
XX	AC	AA47455;	
XX	DT	20-OCT-2000 (first entry)	
XX	DE	Human TANGO 197 coding sequence.	
XX	KW	TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;	
KW	KW	graft versus-host diseases; rheumatoid arthritis; psoriasis;	
KW	KW	inflammatory bowel disease; septic shock; ulcerative colitis;	
KW	KW	Crohn's disease; chronic myelogenous leukemia; cancer; liver	
KW	KW	disease; Hodgkin's disease; osteoarthritis; Lyme's disease;	
KW	KW	cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;	
KW	KW	systemic lupus erythematosus; transgenic animal; diagnosis;	
XX	OS	prognosis; prophylactic; therapeutic; human; ds.	
XX	OS	Homo sapiens.	
XX	FH	Key	Location/Qualifiers
FT	FT	CDS	213..1214
FT	FT	/tag- a	
FT	FT	/product=	TANGO 197
XX	XX	WO200039284-A1.	
XX	XX	06-JUL-2000.	
XX	XX	23-DEC-1999;	99WO-US31025.
XX	XX	30-DEC-1998;	98US-0223546.
XX	XX	(MILL-) MILLENNIUM PHARM INC.	
XX	PI	Holtzman DA;	
XX	XX	WPI; 2000-465743/40.	
DR	DR	P-PSDB; AAB01422.	
XX	PT	Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,	
PT	PT	224 and 239 polypeptides useful for the treatment of asthma, rheumatoid	

PT	arthritis, psoriasis and autoimmune diseases				
XX	Claim 1; Fig 4; 209pp; English.				
XX	Nucleic acids encoding TANGO polypeptides are useful as modulating				
CC	agents for regulating cellular processes like asthma, graft				
CC	versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory				
CC	bowel disease, septic shock, ulcerative colitis, Crohn's disease,				
CC	chronic myelogenous leukemia, cancer, liver disease, Hodgkin's				
CC	disease, osteoarthritis, Lyme's disease, cachexia and autoimmune				
CC	diseases e.g. myasthenia gravis, autoimmune diabetes and systemic				
CC	lupus erythematosus. The nucleic acids are also useful for producing				
CC	transgenic animals and the TANGO polypeptides themselves. Partial				
CC	TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in				
CC	forensic biology, for diagnostic assays, prognostic assays,				
CC	pharmacogenomics and for monitoring clinical trials. TANGO				
CC	polypeptides are suitable for both prophylactic and therapeutic				
CC	methods for treating a subject at risk of a disorder or having a				
CC	disorder associated with aberrant TANGO expression. A wide range				
CC	of cellular disorders can be treated.				
XX	Sequence 2272 BP; 638 A; 554 C; 558 G; 522 T; 0 other;				
SQ					
	Query Match	74.4%;	Score 1052.4;	DB 21;	Length 2272;
	Best Local Similarity	99.9%;	Pred. No. 1.5e-248;		
	Matches 1053;	Conservative	0;	Mismatches	1; Indels 0; Gaps 0;
QY	1	aggaccgcgaggaagggccgcggtgagcgctccctcaggtcgtggcgagttcgcgg	60		
DB	110	aggaccgcgaggaagggccgcggtgagcgctccctcaggtcgtggcgagttcgcgg	169		
QY	61	agcgtgggaaggaagcgcgcctctctcccggtctcccgccatgcccacggcgagcg	120		
DB	170	agcgtgggaaggaagcgcgcctctctcccggtctcccgccatgcccacggcgagcg	229		
QY	121	gagagccctcgcatcggcttccagtgctctcttggccactcgtggtccatctcgc	180		
DB	230	gagagccctcgcatcggcttccagtgctctcttggccactcgtggtccatctcgc	289		
QY	181	cgggcaaggggagcagcagggagatgggggtccagcctcgtacggcgatttgacctga	240		
DB	290	cgggcaaggggagcagcagggagatgggggtccagcctcgtacggcgatttgacctga	349		
QY	241	cttcattttgacaaatcagaagtgctcaccacacacacacacacacacacacacac	300		
DB	350	cttcattttgacaaatcagaagtgctcaccacacacacacacacacacacacacac	409		
QY	301	ggaacagttggtcacaattcattcagccacagttgagaatgctcctttattgtttctc	360		
DB	410	ggaacagttggtcacaattcattcagccacagttgagaatgctcctttattgtttctc	469		
QY	361	caccggaggaacaaccttaataagaaactgacagaagacagaaacacacacacacac	420		
DB	470	caccggaggaacaaccttaataagaaactgacagaagacagaaacacacacacacac	529		
QY	421	agaagaactccagaagttctccagggagagacacttacatgcatgaaggtttgaaag	480		
DB	530	agaagaactccagaagttctccagggagagacacttacatgcatgaaggtttgaaag	589		
QY	481	ggccagtgagcagatttattatgaaaacagacaaggggtacagagcagcagcgtcatc	540		
DB	590	ggccagtgagcagatttattatgaaaacagacaaggggtacagagcagcagcgtcatc	649		
QY	541	tgctttgactgatgagaaactccatgaagatctcttttcttattcagagaggggctaa	600		
DB	650	tgctttgactgatgagaaactccatgaagatctcttttcttattcagagaggggctaa	709		
QY	601	taggtctcagatcttggtgcaattgtttactgtgtgttggaagatttcaatgagac	660		
DB	710	taggtctcagatcttggtgcaattgtttactgtgtgttggaagatttcaatgagac	769		
QY	661	acagctggcccgatttcgcgacagtaaggatcatgtgttcccgtaagacggctttca	720		



```

Db 770 acagctgcccgcgattgscgagcagtaagcatcatgtgtttccggtgaatgcgctttca 829
Qy 721 ggctctgcaaggcatccactcaactcaattttgaagaagtcctgcagtcgaaaattctagcgc 780
Db 830 ggctctgcaaggcatccactcaactcaattttgaagaagtcctgcagtcgaaaattctagcgc 889
Qy 781 tgaaccatccaccatattgcagagagtcatttcaagttgtcgtgagagaaacgcgtt 840
Db 890 tgaaccatccaccatattgcagagagtcatttcaagttgtcgtgagagaaacgcgtt 949
Qy 841 ccgacatcccgcaacgtggacaggtcctctgcagcttcaagatccaatgactcgttcac 900
Db 950 ccgacatcccgcaacgtggacaggtcctctgcagcttcaagatccaatgactcgttcac 1009
Qy 901 actcaatgagaagccctttctgtggaagacacttatttactgtgtccagcgcctattt 960
Db 1010 actcaatgagaagccctttctgtggaagatacttatttactgtgtccagcgcctattt 1069
Qy 961 aaagaagtgtgcatgaaagtcgactccaggtcagcatgaacgatggcctctctttat 1020
Db 1070 aaagaagtgtgcatgaaagtcgactccaggtcagcatgaacgatggcctctctttat 1129
Qy 1021 ctccagttctgtcatcatcacaccacacactgt 1054
Db 1130 ctccagttctgtcatcatcacaccacacactgt 1163

RESULT 4
AAI59918
ID AAI59918 standard; cDNA: 1402 BP.
XX
AC AAI59918;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 3907.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0682191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR P-PSDB; AAM40762.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
```

```

XX Claim 1; SEQ ID NO 3907; 10078pp; English.
PS
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
SQ Sequence 1402 BP; 360 A; 354 C; 399 G; 289 T; 0 other;

Query Match      64.1%; Score 906.8; DB 22; Length 1402;
Best Local Similarity 99.8%; Pred. NO. 7.1e-213;
Matches 908; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aggaccgcgagggaagggccgcgagatggcgctccctcaggggtcgtggcgagatccgcgg 60
Db 206 aggaccgcgagggaagggccgcgagatggcgctccctcaggggtcgtggcgagatccgcgg 265
Qy 61 agcgtggaagagcgagccctcgtctcccgggctgcggccatggccacggcgagcg 120
Db 266 agcgtggaagagcgagccctcgtctcccgggctgcggccatggccacggcgagcg 325
Qy 121 gagagccctcgcagtcggcttcagtgctctctttggccactcgttgctcatctgcgc 180
Db 326 gagagccctcgcagtcggcttcagtgctctctttggccactcgttgctcatctgcgc 385
Qy 181 cgggcaaggggacgcagggagatgggggtccagcctgcctacggcgatggactgta 240
Db 386 cgggcaaggggacgcagggagatgggggtccagcctgcctacggcgatggactgta 445
Qy 241 ctctcatttgacaaatcaggaagtgtgctgcaccacatggaaatgaatctattctgt 300
Db 446 ctctcatttgacaaatcaggaagtgtgctgcaccacatggaaatgaatctattctgt 505
Qy 301 ggaacagttggtcacaattcatcagccacagttgagaatgtccttattgtttctc 360
Db 506 ggaacagttggtcacaattcatcagccacagttgagaatgtccttattgtttctc 565
Qy 361 caccgaggaacaaccttaataagaaactgacagaagacagagaaatccgtcaaggct 420
Db 566 caccgaggaacaaccttaataagaaactgacagaagacagagaaatccgtcaaggct 625
Qy 421 agaagaactccagaagtctctccaggagagacacttacaatgcattgaagattgaaag 480
Db 626 agaagaactccagaagtctctccaggagagacacttacaatgcattgaagattgaaag 685
Qy 481 ggcagtgagcagattttattgaaaacagacaaggggtacagagcagcagcgtcatcat 540
Db 686 ggcagtgagcagattttattgaaaacagacaaggggtacagagcagcagcgtcatcat 745
Qy 541 tgccttgatgagagaaactccatgaagatctcttttctattcagagagggaggttaa 600
Db 746 tgccttgatgagagaaactccatgaagatctcttttctattcagagagggaggttaa 805
Qy 601 taggtctcgagatctgtgtgcaattgttactgtgtgtgtgaaagatttcaatgagac 660
Db 806 taggtctcgagatctgtgtgcaattgttactgtgtgtgtgaaagatttcaatgagac 865
Qy 661 acagctgcccgcgattgscgagcagtaagcatcatgtgtttccggtgaatgcgctttca 720
Db 720 acagctgcccgcgattgscgagcagtaagcatcatgtgtttccggtgaatgcgctttca
```





AAAA479  
ID AAA479 standard; cDNA; 4417 BP.  
AC AAA479;  
XX  
DT 20-OCT-2000 (first entry)  
XX  
DE Murine TANGO 197 coding sequence.  
XX  
KW TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;  
KW graft versus-host diseases; rheumatoid arthritis; psoriasis;  
KW inflammatory bowel disease; septic shock; ulcerative colitis;  
KW Crohn's disease; chronic myelogenous leukemia; cancer; liver  
KW disease; Hodgkin's disease; osteoarthritis; Lyme's disease;  
KW cachexia; autoimmune disease; myasthenia gravis; autoimmune diabetes;  
KW systemic lupus erythematosus; transgenic animal; diagnosis;  
KW prognosis; prophylactic; therapeutic; mouse; ds.  
XX  
OS Mus musculus.  
XX  
FH Key Location/Qualifiers  
CDS 3..1148  
FT /\*tag= a  
FT /product= TANGO 197  
FT  
XX WO200039284-A1.  
XX  
XX 06-JUL-2000.  
XX  
XX 23-DEC-1999; 99WO-US31025.  
XX  
XX 30-DEC-1998; 98US-0223546.  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
XX  
XX Holtzman DA;  
XX  
XX WPI; 2000-465743/40.  
XX P-PSDB; AAB01428.  
XX  
XX Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213,  
XX 224 and 239 polypeptides useful for the treatment of asthma, rheumatoid  
XX arthritis, psoriasis and autoimmune diseases  
XX  
XX Claim 1; Fig 27; 209pp; English.  
XX  
XX Nucleic acids encoding TANGO polypeptides are useful as modulating  
XX agents for regulating cellular processes like asthma, graft  
XX versus-host diseases, rheumatoid arthritis, psoriasis, inflammatory  
XX bowel disease, septic shock, ulcerative colitis, Crohn's disease,  
XX chronic myelogenous leukemia, cancer, liver disease, Hodgkin's  
XX disease, osteoarthritis, Lyme's disease, cachexia and autoimmune  
XX diseases e.g. myasthenia gravis, autoimmune diabetes and systemic  
XX lupus erythematosus. The nucleic acids are also useful for producing  
XX transgenic animals and the TANGO polypeptides themselves. Partial  
XX TANGO-128, 140, 197, 212, 213, 224, 239 sequences are useful in  
XX forensic biology, for diagnostic assays, prognostic assays,  
XX pharmacogenomics and for monitoring clinical trials. TANGO  
XX polypeptides are suitable for both prophylactic and therapeutic  
XX methods for treating a subject at risk of a disorder or having a  
XX disorder associated with aberrant TANGO expression. A wide range  
XX of cellular disorders can be treated.  
XX  
XX Sequence 4417 BP; 1214 A; 1115 C; 994 G; 1094 T; 0 other;  
XX  
Query Match 32.1%; Score 453.2; DB 21; Length 4417;  
Best Local Similarity 90.1%; Pred. No. 3.6e-101;  
Matches 485; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
XX  
QY 660 cacagctggccggattgcggacagatcattgtttccctggaatcagcgctttc 719  
DB 10 ctcatgttgctgattgcagacagtaaggaccacgtgtttcctgtgaacgagcttcc 69

QY 720 aggcctctgaagcattccactcaattttgaagaagtccctgcacatcgaattctctgacg 779  
DB 70 aggcctccaagcattatccactcaattttaaagaaatccctgcacatcgaattctgagg 129  
QY 780 ctgaaccatccacatattgcaggagagtcattcaagtctgcgtgagaggaacggct 839  
DB 130 ctgaaccatccacatctgcgcggagagtcctttcaagtgcgtaaggaggaatggct 189  
QY 840 tcgacatgccgcaacgtgacagaggtcctctgcagcttcaagatcaatgactcggctca 899  
DB 190 tcgacatgccgcaatgtgacagggcctctgcagcttcaaaaatcaatgactcagtc 249  
QY 900 cactcaatgagagccctttctgtggaagacacttatttactgtgtccagcgctctatct 959  
DB 250 cgctcaatgagagccctttctgtggaagacacttatttctgtgtgccagcacaatct 309  
QY 960 taaaagaagtggcatgaaagctgcactccaggtcagatcagatgaacgatgctctcttta 1019  
DB 310 tgaagaagtggcatgaaagctgcactgcaggtcagatgaacgatgctctcttta 369  
QY 1020 tctcagttctgcactcaccacacacactgttctgcaggttccatctctggccatcg 1079  
DB 370 tctcagttctgcactcaccacacacactgttctgcaggttccatctctggcgattg 429  
QY 1080 cctctgctgctctgttctctctctgctctcctctctctctctctctctctctctct 1139  
DB 430 ctctgctgctctctctctctctctctctctctctctctctctctctctctctctct 489  
QY 1140 gctgacatgattatcaagaggtccctccacccctcgcgagagagtgaggaataa 1197  
DB 490 gctgacagtgatcatcaagaggtccctccacccctcgcgagagagtgaggaataa 547  
RESULT 8  
AAH06188  
ID AAH06188 standard; cDNA; 464 BP.  
XX  
AC AAH06188;  
XX  
XX 26-JUN-2001 (first entry)  
XX  
XX Human cDNA clone (5'-primer) SEQ ID NO:3023.  
XX  
XX Human, primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX  
XX EP1074617-A2.  
XX  
XX 07-FEB-2001.  
XX  
XX 28-JUL-2000; 2000EP-0116126.  
XX  
XX 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300253.  
XX 11-JAN-2000; 2000JP-0118776.  
XX 02-MAY-2000; 2000JP-0183767.  
XX 09-JUN-2000; 2000JP-0241899.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
XX WPI; 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
XX full-length cDNAs defined in the specification, and for the detection  
XX and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -  
XX  
XX Claim 1; SEQ ID 3023; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesising polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX Sequence 464 BP; 101 A; 122 C; 141 G; 90 T; 10 other;

Query Match 26.0%; Score 368; DB 22; Length 464;  
 Best Local Similarity 93.9%; Pred. No. 1.1e-80;  
 Matches 388; Conservative 0; Mismatches 24; Indels 1; Gaps 1;  
 Qy 12 ggaagggccgcgagtgccgctccctcagggctgctgaggtcgtgcgaggtcgcgagcgtggaag 71  
 Db 53 ggaagggccgcgagtgccgctccctcagggctgctgaggtcgtgcgaggtcgcgagcgtggaag 112  
 Qy 72 gaggcgaccctgctctcccccgggtgcggccatggccacgcgagcgagagccctcg 131  
 Db 113 gaggcgaccctgctctcccccgggtgcggccatggccacgcgagcgagagccctcg 172  
 Qy 132 gcatcggttccagtggtctctcttctggccactctggtctcatctatctgcgccgggcaagg 191  
 Db 173 gcatcggttccagtggtctctc-acggccactctggtctcatctatctgcgccgggcaagg 231  
 Qy 192 gacgagagagtgagggttcacgctctacgcgaggttgacctgactctatttgg 251  
 Db 232 gacgagagagtgagggttcacnctctacgcgaggttgacctgactctatttgg 291  
 Qy 252 acaaatcaggagtgctgctgcaccactggaatgaaatctattcttggaaacagttgg 311  
 Db 292 acaaatcaggagtgctgctgcaccactggaatgaaatctattcttggaaacagttgg 351  
 Qy 312 ctcaaaattcaccagccacagtgagaaatgctcttattgtttctccaccggagaa 371  
 Db 352 ctcaacnaattcatcaaccacagttgaaatgctcttattgtttctccaccagaa 411  
 Qy 372 caacttaataaactgacagaagacagaaacaaatccgtcaagccctaga 424  
 Db 412 caactttatgaactgacggaanacagaaaccantccctcnggctcnaa 464

RESULT 9  
 AAL19043  
 ID AAL19043 standard; cDNA; 356 BP.  
 XX AAL19043;  
 AC AAL19043;  
 XX  
 DT 07-DEC-2001 (first entry)  
 XX Human breast cancer expressed polynucleotide 11500.  
 DE Human; breast cancer; cell marker; cytostatic; ss.  
 KW Homo sapiens.  
 OS

XX PN WO200151628-A2.  
 XX PD 19-JUL-2001.  
 XX PF 10-JAN-2001; 2001WO-US00798.  
 XX PR 14-JAN-2000; 2000US-0176077.  
 PR 14-MAR-2000; 2000US-0189167.  
 PR 24-MAR-2000; 2000US-0192099.  
 PR 29-MAR-2000; 2000US-0193480.  
 PR 15-MAY-2000; 2000US-0205230.  
 PR 09-JUN-2000; 2000US-0211315.  
 PR 25-JUL-2000; 2000US-0220534.  
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX Lillie J, Xu Y, Wang Y, Steinmann K;  
 PI WPI; 2001-451856/48.  
 XX New peptide useful as a marker for the diagnosis of breast cancer -  
 PT Claim 1; Page 2056; 3695pp; English.  
 XX The invention relates to human breast cancer expressed polynucleotides  
 CC (AAL07544-AAL26789) and methods of assessing whether a patient is  
 CC afflicted with breast cancer by examining the correlation between the  
 CC expression of certain markers and the cancerous state of breast cells.  
 CC The polynucleotides and encoded polypeptides are potential markers for  
 CC detecting, diagnosing, monitoring, characterising treating and  
 CC potentially preventing breast cancer. The polynucleotides and encoded  
 CC polypeptides are also useful for isolating compounds with cytostatic  
 CC activity.  
 XX SQ Sequence 356 BP; 91 A; 78 C; 95 G; 92 T; 0 other;  
 Query Match 24.7%; Score 349; DB 22; Length 356;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-76;  
 Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 516 ggtacagacagccagcgtcatctattgttgaactgaggaactccatgaagatctct 575  
 Db 8 ggtacagacagccagcgtcatctattgttgaactgaggaactccatgaagatctct 67  
 Qy 576 tttctattcagagaggaggttaaggtctcagatcttgggtgcaattgtttactgtg 635  
 Db 68 tttctattcagagaggaggttaaggtctcagatcttgggtgcaattgtttactgtg 127  
 Qy 636 ttggtgtgaagatttcaatgagacacagcgtgcccggattcggacagtaagatcatg 695  
 Db 128 ttggtgtgaagatttcaatgagacacagcgtgcccggattcggacagtaagatcatg 187  
 Qy 696 ttttcccgctgaatgacgcttccaggctctcgaaggatccatccactcaatttgaaga 755  
 Db 188 ttttcccgctgaatgacgcttccaggctctcgaaggatccatccactcaatttgaaga 247  
 Qy 756 agtctcgtcatcgaaattctagcagctgaacctccaccatattgtgcaggagagtcatttc 815  
 Db 248 agtctcgtcatcgaaattctagcagctgaacctccaccatattgtgcaggagagtcatttc 307  
 Qy 816 aagtgtcgtgagagagaaacggcttccgacatgcccgcaacgtgacag 864  
 Db 308 aagtgtcgtgagagagaaacggcttccgacatgcccgcaacgtgacag 356  
 RESULT 10  
 AAL08133  
 ID AAL08133 standard; cDNA; 355 BP.  
 XX AAL08133;  
 AC AAL08133;  
 XX

```
DT 07-DEC-2001 (first entry)
XX Human breast cancer expressed polynucleotide 590.
XX Human; breast cancer; cell marker; cytostatic; ss.
XX Homo sapiens.
XX WO200151628-A2.
XX 19-JUL-2001.
XX 10-JAN-2001; 2001WO-US00798.
XX 14-JAN-2000; 2000US-0176077.
XX 14-MAR-2000; 2000US-0189167.
XX 24-MAR-2000; 2000US-0192099.
XX 29-MAR-2000; 2000US-0193480.
XX 15-MAY-2000; 2000US-0205230.
XX 09-JUN-2000; 2000US-0211315.
XX 25-JUL-2000; 2000US-0220534.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Lillie J, Xu Y, Wang Y, Steinmann K;
XX WPI; 2001-451856/48.
XX New peptide useful as a marker for the diagnosis of breast cancer -
XX Claim 1; Page 189; 3695pp; English.
XX The invention relates to human breast cancer expressed polynucleotides
XX (AAL07544-AAL26789) and methods of assessing whether a patient is
XX afflicted with breast cancer by examining the correlation between the
XX expression of certain markers and the cancerous state of breast cells.
XX The polynucleotides and encoded polypeptides are potential markers for
XX detecting, diagnosing, monitoring, characterising treating and
XX potentially preventing breast cancer. The polynucleotides and encoded
XX polypeptides are also useful for isolating compounds with cytostatic
XX activity.
XX Sequence 355 BP; 91 A; 78 C; 94 G; 92 T; 0 other;

Query Match 24.6%; Score 348; DB 22; Length 355;
Best Local Similarity 100.0%; Pred. No. 8.3e-76;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 ggtacagacagccagcgtcatcattgtttgactgattggagaaactccatgaagatctct 575
DB 8 ggtacagacagccagcgtcatcattgtttgactgattggagaaactccatgaagatctct 67
QY 576 tttttattcagagaggagcctaataagttctcgagattcgttgcaattgtttactgtg 635
DB 68 tttttattcagagaggagcctaataagttctcgagattcgttgcaattgtttactgtg 127
QY 636 ttggtgtgaaagatttcaatgagacacagctggcccgattgctgagacagtagaatcatg 695
DB 128 ttggtgtgaaagatttcaatgagacacagctggcccgattgctgagacagtagaatcatg 187
QY 696 tttttccgtgaatgacggcttccaggtctcgaaggatcatccactcaattttgaaga 755
DB 188 tttttccgtgaatgacggcttccaggtctcgaaggatcatccactcaattttgaaga 247
QY 756 agtctctgcatcgaattcttagcagctgaaccatccaccatattgtgcaggagagtcatttc 815
DB 248 agtctctgcatcgaattcttagcagctgaaccatccaccatattgtgcaggagagtcatttc 307
QY 816 aagtgtcgtgagagaaacggcttccgacatgcccgcaacgtaggaca 863
DB 308 aagtgtcgtgagagaaacggcttccgacatgcccgcaacgtaggaca 355
```

---

```
RESULT 11
AAL25878
ID AAL25878 standard; cDNA; 346 BP.
XX
XX AAL25878;
XX
XX 07-DEC-2001 (first entry)
XX Human breast cancer expressed polynucleotide 18335.
XX Human; breast cancer; cell marker; cytostatic; ss.
XX Homo sapiens.
XX WO200151628-A2.
XX 19-JUL-2001.
XX 10-JAN-2001; 2001WO-US00798.
XX 14-JAN-2000; 2000US-0176077.
XX 14-MAR-2000; 2000US-0189167.
XX 24-MAR-2000; 2000US-0192099.
XX 29-MAR-2000; 2000US-0193480.
XX 15-MAY-2000; 2000US-0205230.
XX 09-JUN-2000; 2000US-0211315.
XX 25-JUL-2000; 2000US-0220534.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Lillie J, Xu Y, Wang Y, Steinmann K;
XX WPI; 2001-451856/48.
XX New peptide useful as a marker for the diagnosis of breast cancer -
XX Claim 1; Page 3388; 3695pp; English.
XX The invention relates to human breast cancer expressed polynucleotides
XX (AAL07544-AAL26789) and methods of assessing whether a patient is
XX afflicted with breast cancer by examining the correlation between the
XX expression of certain markers and the cancerous state of breast cells.
XX The polynucleotides and encoded polypeptides are potential markers for
XX detecting, diagnosing, monitoring, characterising treating and
XX potentially preventing breast cancer. The polynucleotides and encoded
XX polypeptides are also useful for isolating compounds with cytostatic
XX activity.
XX Sequence 346 BP; 90 A; 75 C; 90 G; 91 T; 0 other;

Query Match 24.5%; Score 346; DB 22; Length 346;
Best Local Similarity 100.0%; Pred. No. 2.6e-75;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 519 acagagacagcagcgtcatcattgtttgactgattggagaaactccatgaagatctcttt 578
DB 1 acagagacagcagcgtcatcattgtttgactgattggagaaactccatgaagatctcttt 60
QY 579 tctattcagagaggaggcctaataagttctcgagattcgttgcaattgtttactgtgtg 638
DB 61 tctattcagagaggaggcctaataagttctcgagattcgttgcaattgtttactgtgtg 120
QY 639 gtgtgaaagatttcaatgagacacagctggcccgattgctgagacagtagaatcatgtgt 698
DB 121 gtgtgaaagatttcaatgagacacagctggcccgattgctgagacagtagaatcatgtgt 180
QY 699 ttcocgtgaatgacggcttccaggtctcgaaggatcatccactcaattttgagaagt 758
DB 181 ttcocgtgaatgacggcttccaggtctcgaaggatcatccactcaattttgagaagt 240
QY 759 cctgcagaaattctagcagctgaaccatccacatattgtgcaggagagatcattcaag 818
```

Db 241 cctgcagaaattctagcagtaaacatccaccatattgtaggagagcattccaag 300  
 Qy 819 ttgtcgtgagaggaacggctccgacatgcccgaacgtggacag 864  
 Db 301 ttgtcgtgagaggaacggctccgacatgcccgaacgtggacag 346

RESULT 12  
 AAA75149  
 ID AAA75149 standard; cDNA: 3677 BP.  
 XX  
 AC AAA75149;  
 XX  
 XX 15-JAN-2001 (first entry)  
 XX  
 DE cDNA encoding a human TANGO 216 polypeptide.  
 XX  
 KW TANGO 266; TANGO 216; TANGO 262; TANGO 267;  
 KW cellular proliferation; cellular differentiation; cellular adhesion;  
 KW von Willebrand factor-associated disorder; cell trafficking; cancer;  
 KW hematopoietic associated disease; atelectasis; pulmonary congestion;  
 KW oedema; emphysema; chronic bronchitis; bronchial asthma; bronchiectasis;  
 KW intestinal disorder; spleen associated disease; renal disorder;  
 KW cardiovascular disorder; ischemic heart disease; hydrocephalus;  
 KW brain herniation; iatrogenic disease; inflammation; meningitis;  
 KW Alzheimer's Disease; cerebral toxoplasmosis; Parkinson's disease;  
 KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 307..173  
 FT /\*tag= a  
 FT /product= "TANGO 216"  
 FT sig\_peptide 307..405  
 FT /\*tag= b  
 FT mat\_peptide 406..173  
 FT /\*tag= c

XX WO200052022-A1.  
 PN 08-SEP-2000.  
 XX  
 XX 01-MAR-2000; 2000WO-US05226.  
 PF  
 XX 01-MAR-1999; 99US-0122458.  
 PR  
 XX (MILL-) MILLENNIUM PHARM INC.  
 PA  
 PI Barnes TM, Holtzman DA, Sharp JD, Fraser CC;  
 XX  
 XX WPI: 2000-579269/54.  
 DR P-PSDB; AAB18447.  
 XX  
 XX Novel human and murine secreted proteins designated TANGO 216, 261,  
 PT 262, 266 and 267 useful as modulating agents of cellular processes,  
 PT e.g. for treating cancer -  
 XX  
 PS Claim 2; Fig 1A-C; 175pp; English.  
 XX  
 CC The present sequence encodes a human TANGO 216 polypeptide. The  
 CC specification also describes TANGO 266, TANGO 261, TANGO 262, and  
 CC TANGO 267. The TANGO polypeptides can be used to modulate cellular  
 CC proliferation, modulate cellular differentiation and/or modulate  
 CC cellular adhesion. The proteins can be used to treat any von Willebrand  
 CC factor-associated disorder, regulate extracellular matrix structuring,  
 CC cellular adhesion, and cell trafficking and/or migration, modulate  
 CC cellular interactions, modulate cell adhesion in proliferative  
 CC disorders, such as cancer, modulate the proliferation, differentiation,  
 CC and/or function of cells that appear in the bone marrow, and leukocytes,  
 CC treat bone marrow, blood and hematopoietic associated diseases and  
 CC disorders, atelectasis, pulmonary congestion or oedema, emphysema,

CC chronic bronchitis, bronchial asthma and bronchiectasis, intestinal  
 CC disorders, spleen associated diseases, modulate renal disorders, treat  
 CC cardiovascular disorders such as ischemic heart disease, modulate the  
 CC proliferation, differentiation, and/or function of bone and cartilage  
 CC cells and to treat bone and/or cartilage associated diseases or  
 CC disorder. They may also be used to treat disorders associated with the  
 CC ovaries, cerebral oedema, hydrocephalus, brain herniations, iatrogenic  
 CC disease, inflammations, bacterial and viral meningitis, Alzheimer's  
 CC Disease, cerebral toxoplasmosis, Parkinson's disease, multiple sclerosis,  
 CC brain cancers, hydrocephalus and encephalitis, and treat hepatic  
 CC disorders.  
 XX  
 SQ Sequence 3677 BP; 1055 A; 776 C; 881 G; 965 T; 0 Other;  
 XX

Query Match 24.5%; Score 345.8; DB 21; Length 3677;  
 Best Local Similarity 57.1%; Pred. No. 8.2e-75;  
 Matches 672; Conservative 0; Mismatches 497; Indels 8; Gaps 2;

Qy 4 accgcgaggaagggcccgatggcgctccctcgaggtcggtggcgagttcgcgagc 63  
 Db 209 acccagagtgctgctggcggtgactcccgccacctttgcacccctcctgagcttagggga 268  
 Qy 64 gtgggaagagcgaccctctctcccggtctcgccgcatggccacgcgcgagcgag 123  
 Db 269 ctgcgagcgagggagctctcagggcccgccgcagc--atggtggcgagcggtccccc 326  
 Qy 124 agcctcgagctggctccagtggtctctttggcactctgtgtcactctgcgcgg 183  
 Db 327 ggcgcgagcccgagtgctgtcccggtgtggtgtgtgtgtgtgtgtgtgtgtgtgt 386  
 Qy 184 gcaaggggacgagggaggtgggtgtccagctgtcacggtgtgtgtgtgtgtgtgtgtgt 243  
 Db 387 cgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 446  
 Qy 244 cattttggacaaatcaggaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 303  
 Db 447 cgtcgtgacaaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 506  
 Qy 304 acagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 363  
 Db 507 gcaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 566  
 Qy 364 cggaggaacaccttaataatgaaactgacagagacagagacaaacccgtcaagcgctaga 423  
 Db 567 tcaagcaactattattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 626  
 Qy 424 agactccagaaattctgtccagggagagacacttcatgtcatgaggaaggtgtgtgtgtgt 483  
 Db 627 gatttaaaacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 686  
 Qy 484 cagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 543  
 Db 687 gaatgaacaaat-----tcagaaagcagggaggtgtgtgtgtgtgtgtgtgtgtgtgt 740  
 Qy 544 tttaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 603  
 Db 741 tctgacagatggcaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 800  
 Qy 604 gtctcgagatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 663  
 Db 801 atccaggtcacttgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 860  
 Qy 664 gctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 723  
 Db 861 gcttgaagaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 920  
 Qy 724 tctgcaagcagcatccactcaattttgaagaagtcctgtcatcgaatctctagcagctga 783  
 Db 921 tcttaaggaataataattctactagctcagtcagtcagtcagtcagtcagtcagtcagtc 980  
 Qy 784 accatccaccatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 843





Db 741 tctgacagatggcaagtggacggtctgtggtccatcatatgacagagaaggcaagat 800  
 QY 604 gtctcgagatcttggtgcaattgtttactgtgttggtgtaaaagatttcaatgagacaca 663  
 Db 801 atccaggtcacttggggtctgtgtttattgtgttgggtctgtgtgttgaacaagacaca 860  
 QY 664 gctggcccggtggtgacagagtaaggatcatgtgtttcccggtgaatgacgcttccaggc 723  
 Db 861 gctgaaagaattgctgattcccaaggagcaagttttccctggtcaaaagtggatttcaggc 920  
 QY 724 tctgcaaggcatcatccactcaattttgaagaagtcgtcgtcatcgaaattcttagcagtga 783  
 Db 921 tcttaaggaataaataattctactacagctcagctcagctcagctcagctcagctcagct 980  
 QY 784 accatccaccatattgacagagatcatcttcaagtgtgtgtgagagaaacggtctccg 843  
 Db 981 gcccctcaagtgtctgtgtgggggggaatttccagattgtcttcaagtggaaaggattcat 1040  
 QY 844 acatgcccgaacgtggacaggggtctctctgcagcttccaagatcaatgactcgtgcacact 903  
 Db 1041 gctgggagtcggaatggcagtggtctctgcactacactgtaaatgaacatacaac 1100  
 QY 904 caatgaagaagcccttctgtggaagacacttatttactgtgtccagcgcctattctaa 963  
 Db 1101 gagtgtaaaaccagtaagtgtacagcttaattctctgtgtgtgtgtgtgtgtgtgtgtgt 1160  
 QY 964 agaagtggcatgaaagtcgactccaggtcagcatgaacagatggcctctcttttatctc 1023  
 Db 1161 taaagctggagaaactcttgatgtttcagtgagctttaa199agggaaactctgcatttc 1220  
 QY 1024 cagttctgtcatcatcaccacacacactgttctgacggtttccatctctggtccatgcct 1083  
 Db 1221 aggatcataattgtcacagccacagaatgtctaaaggatgcagccatcattgttat 1280  
 QY 1084 gctgactctgtctgtctagccctggtctctctctctgtgtgtgtgtgtgtgtgtgtgtgt 1143  
 Db 1281 ttgtgttactgtactctctggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1340  
 QY 1144 cactgtgattatcaagaggtccctccaccctctgcc 1180  
 Db 1341 caaagtgttataagagatctccaccaccaccctcc 1377

RESULT 14  
 AAA75158  
 ID AAA75158 standard; cDNA; 3677 BP.  
 XX AC AAA75158;  
 XX DT 15-JAN-2001 (first entry)  
 XX DE cDNA clone encoding a human TANGO 216 polypeptide.  
 XX KW TANGO 266; TANGO 216;; TANGO 261; TANGO 262; TANGO 267;  
 KW cellular proliferation; cellular differentiation; cellular adhesion;  
 KW von Willebrand factor-associated disorder; cell trafficking; cancer;  
 KW hematopoietic associated disease; atelectasis; pulmonary congestion;  
 KW oedema; emphysema; chronic bronchitis; atelectasis; bronchial asthma; bronchiectasis;  
 KW intestinal disorder; spleen associated disease; renal disorder;  
 KW cardiovascular disorder; ischemic heart disease; hydrocephalus;  
 KW brain herniation; atrogenic disease; inflammation; meningitis;  
 KW Alzheimer's Disease; cerebral toxoplasmosis; Parkinson's disease;  
 KW multiple sclerosis; hydrocephalus; encephalitis; hepatic disorder; ss.  
 XX OS Homo sapiens.  
 XX FH Key Location/Qualifiers  
 CDS 307..1773  
 FT FT  
 FT /\*tag- a  
 FT /product- "TANGO 216"  
 XX PN WO200052022-A1.  
 XX

PD 08-SEP-2000.  
 XX 01-MAR-2000; 2000WO-USO5226.  
 XX 01-MAR-1999; 99US-0122458.  
 XX (MILL-) MILLENNIUM PHARM INC.  
 XX Barnes TM, Holtzman DA, Sharp JD, Fraser CC;  
 XX WPI: 2000-579269/54.  
 XX P-PSDB: AAB18456.  
 XX Novel human and murine secreted proteins designated TANGO 216, 261,  
 PT 262, 266 and 267 useful as modulating agents of cellular processes,  
 PT e.g. for treating cancer -  
 XX Disclosure; Page -: 175pp; English.  
 XX AAA75157-59 encode human TANGO 216 proteins. The specification also  
 CC describes TANGO 266, TANGO 261, TANGO 262, and TANGO 267. The TANGO  
 CC polypeptides can be used to modulate cellular proliferation, modulate  
 CC cellular differentiation and/or modulate cellular adhesion. The  
 CC proteins can be used to treat any von Willebrand factor-associated  
 CC disorder, regulate extracellular matrix structuring, cellular adhesion,  
 CC and cell trafficking and/or migration, modulate cellular interactions,  
 CC modulate cell adhesion in proliferative disorders, such as cancer,  
 CC modulate the proliferation, differentiation, and/or function of cells  
 CC that appear in the bone marrow, and leukocytes, treat bone marrow, blood  
 CC and hematopoietic associated diseases and disorders, atelectasis,  
 CC pulmonary congestion or oedema, emphysema, chronic bronchitis, bronchial  
 CC asthma and bronchiectasis, intestinal disorders, spleen associated  
 CC diseases, modulate renal disorders, treat cardiovascular disorders such  
 CC as ischemic heart disease, modulate the proliferation, differentiation,  
 CC and/or function of bone and cartilage cells and to treat bone and/or  
 CC cartilage associated diseases or disorder. They may also be used to  
 CC treat disorders associated with the ovaries, and cerebral oedema,  
 CC hydrocephalus, brain herniations, iatrogenic disease, inflammations,  
 CC bacterial and viral meningitis, Alzheimer's Disease, cerebral  
 CC toxoplasmosis, Parkinson's disease, multiple sclerosis, brain cancers,  
 CC hydrocephalus and encephalitis, and treat hepatic disorders.  
 CC note: the present sequence does not appear in the specification; it was  
 CC created using information provided.  
 XX SQ Sequence 3677 BP; 1055 A; 777 C; 880 G; 965 T; 0 other;

Query Match 24.3%; Score 344.2; DB 21; Length 3677;  
 Best Local Similarity 57.0%; Pred. No. 2e-74;  
 Matches 671; Conservative 0; Mismatches 498; Indels 8; Gaps 2;  
 QY 4 acccgaggaagggccgagatggcgctccctcctcaggggtcgtggcgagttcgcggagc 63  
 Db 209 acccagagtcgtgcccgggtgactcccgccacatttgcacccctctcgtagcttagggga 268  
 QY 64 gtgggaagagcgacccctctctcccgggtcgtggccatggccacgcgagcgag 123  
 Db 269 ctggagcgaggagagtcctcagggcccccgcgcagg--atggtggcgagcggtcccc 326  
 QY 124 agccctcgcatcggtctccagtggtctcttggccactctgtgtcatctgcgcgg 183  
 Db 327 ggcgcgagcccgagctggtgttcccggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 386  
 QY 184 gcaaggggagcgaggagatgggggtccagctctcagcggattgacgttactt 243  
 Db 387 cgggggctgctgcgcccaggaccctcctcagagagccttctgacatctt 446  
 QY 244 catttgacaaatcaggaaagtgtgctgacacactggaaatctattcttggga 303  
 Db 447 cgtctggacaagtctggagtggtggcaataactggaattgaaatttataatttcgtaca 506  
 QY 304 acagttggtcacaaatcatcagccacagttgagaatgtccttatttttttccac 363



Search completed: August 9, 2002, 00:30:14  
Job time: 8026 sec

D	b	269	ctgcgagcgggaggtctctcaggccccggccgcaggg--atgggtggcgagcgggtccccc	326
Q	y	124	agccctcggcatcgcgcttccagtggtctctttggccacactctggtctgctatctgcgcgg	183
D	b	327	ggccgcagcccgagagctggctgttcccgggctgtggctgttgggtctcagcggctcc	386
Q	y	184	gcaagggggacgcaggaggatcgggggtccagccgtgtacgcgggatttgacctgtactt	243
D	b	387	cggggggcgtcgtcgcccgagcagccctctctgcagaaagagccttggatctctactt	446
Q	y	244	cattttggacaaatcaggaagtgtctgcaccactggaatgaaatctattacttttggga	303
D	b	447	cgctccggcaagctctgggagtggtggcaaaataactggattgacaattcataattctgtaca	506
Q	y	304	acagtttggctcacaattcatcagcccacagtttgagaatgtcccttattgttttccac	363
D	b	507	gcaactgcggagagattgtgagccctgaaatgagattatcttcatgtgtttcttc	566
Q	y	364	ccgaggaacaaaccttaatgaactgcagaaagacagagaacaaataccgtcaagccctaga	423
D	b	567	tcaagcaactattatttggcattaaactggagacagaggcaaaatcagtaaaagcttggga	626
Q	y	424	agaactccagaagaattctgcaggagggagacacittatcatgcatgaaagatttgaaagggc	483
D	b	627	ggattcaaaacgtgttgctcagtgaggagacatatatccatgaaggagtaaaagctagc	686
Q	y	484	cagtgagcagatttatttgaaaacagacaaggggtacaggacagccagcgtcatcattgc	543
D	b	687	gaatgacaaaaat-----tcagaagaacaggaggccttgaaaacctccagtatcataattgc	740
Q	y	544	tttgactgtggagaactccatgaagatctctttttcttcttcagcagaggaggctaatag	603
D	b	741	tctgcagatggcaattggacggtctgtggtgccatcatatgcagagaagagcgaagat	800
Q	y	604	gtctcagatcttggtgcaattgttttactgtgtgttggtgaaagatttcaatgagacaca	663
D	b	801	atccaggtcacctggggcagtggttattgtgtgtgtgctctgtatttgaacaagcaca	860
Q	y	664	gctggcccgattgcggacagtaagagatcatgtgttcccggtgaatgacggcttcaggc	723
D	b	861	gcttgaaagaattgctgattccaaggagcaggtttccctgtcaaaaggtggatttcaggc	920
Q	y	724	tctgcaggcatcatccactcaatttgaagaagctctgcatacgaaattctagcagctga	783
D	b	921	tcttaaggaaataataattctatactagctcagtcagtcagtactgaaatcctagaattgca	980
Q	y	784	accatccaccatatgtgcaggagagtcatttcaagttgtctgtagagaagaacggcttcog	843
D	b	981	gcctcgaagtgtctgtgtggggagaaattcagattgtcttcaagtggagaggattcat	1040
Q	y	844	acatgcccgcaacgtggacagggctcctgcagctccaagatcaatgactcggtcacact	903
D	b	1041	gctgggcagtcggaaatggcagtggtctctgcacttacactgaaatgaacatatacaac	1100
Q	y	904	caatgagaagccctttctgttgaagacacatttattactgtgtccagcgcctactttaaa	963
D	b	1101	gagtgtaaaaccagtaagtgtaagcttaattctatgttctgtcctgcacctatcctgaa	1160
Q	y	964	agaagttggcagtgaagagctgcactccagtcagaatgaacgatggcctctcttttatctc	1023
D	b	1161	taaagctggagaacctctgtatttcagtgagcgtttaaagtgaggaaaaattgtcatttc	1220
Q	y	1024	cagttctgtcatcatcacaccacacactgtttcgaaggttccatcctggccatcgccct	1083
D	b	1221	aggaatcaattgtcacagccacagaaatgttctaacyggatcgcagccatcatgtttat	1280
Q	y	1084	gctgatacctgttccctctagcctggctctctctctgtgtgttcttgcccctctgctg	1143
D	b	1281	tttgggttactgctactcctgggaatcggtttgatgtgtgtgttttggccctttgtcg	1340
Q	y	1144	cactgtgattatacgaagggtccctccaccocctgcc	1180
D	b	1341	caaaagtgttatlaaagatccctccaccacacccccc	1377



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2002, 10:46:58 ; Search time 31.3 Seconds  
(without alignments)  
364.929 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_27\_321

Perfect score: 1536

Sequence: 1 GCGGREDGPGACGGGFLY.....GLSFSSVITTTTHCSGDS 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1536	100.0	564	1 ATR_HUMAN	Q9h6x2 homo sapien
2	1519	98.9	562	1 ATR_MOUSE	Q9cz52 mus musculus
3	534.5	34.8	386	1 CMG2_HUMAN	P58335 homo sapien
4	156	10.2	1162	1 ITAD_HUMAN	Q13349 homo sapien
5	145.5	9.5	1152	1 ITAM_HUMAN	P11215 homo sapien
6	143.5	9.3	1163	1 ITAX_HUMAN	P20702 homo sapien
7	143	9.3	3124	1 CA1C_CHICK	P13944 gallus gall
8	141	9.2	1153	1 ITAM_MOUSE	P05555 mus musculus
9	139	9.0	3067	1 CA1C_MOUSE	Q60847 mus musculus
10	131	8.5	760	1 CO2_MOUSE	P21180 mus musculus
11	131	8.5	3063	1 CA1C_HUMAN	Q99715 homo sapien
12	130	8.5	1888	1 CA1E_CHICK	P32018 gallus gall
13	128	8.3	3051	1 YNX3_CAEEL	P34576 caenorhabdi
14	118	7.7	929	1 CA1C_NOTVI	Q91145 notophthalm
15	113.5	7.4	496	1 CAMA_HUMAN	P21941 homo sapien
16	113.5	7.4	764	1 CFAB_HUMAN	P00751 homo sapien
17	113.5	7.4	956	1 MTN2_HUMAN	O00339 homo sapien
18	111.5	7.3	452	1 MTN3_CHICK	O42401 gallus gall
19	111.5	7.3	2844	1 CA17_HUMAN	Q02388 homo sapien
20	110.5	7.2	761	1 CFAB_MOUSE	P04186 mus musculus
21	109.5	7.1	1029	1 CA26_MOUSE	Q02788 mus musculus
22	108.5	7.1	500	1 CAMA_MOUSE	P51942 mus musculus
23	108.5	7.1	956	1 MTN2_MOUSE	O08746 mus musculus
24	107.5	7.0	493	1 CAMA_CHICK	P05099 gallus gall
25	107	7.0	1189	1 ITAH_HUMAN	Q9ukx3 homo sapien
26	104.5	6.8	624	1 MTN4_MOUSE	O89029 mus musculus
27	102	6.6	486	1 MTN3_HUMAN	O15232 homo sapien
28	99.5	6.5	597	1 ROP_DROME	Q07327 drosophila
29	99	6.4	559	1 TRAP_PLAFA	P16893 plasmodium
30	99	6.4	619	1 MTN4_HUMAN	O95460 homo sapien
31	99	6.4	1018	1 CA26_HUMAN	P12110 homo sapien
32	98	6.4	3176	1 CA36_HUMAN	P12111 homo sapien
33	97.5	6.3	642	1 DNAK_NEIMB	Q9K0n4 neisseria m

## RESULT 1

ID	ATR_HUMAN	STANDARD;	PRT;	564 AA.
AC	Q9H6X2; OSNVP3;			
DT	01-MAR-2002 (Rel. 41, Created)			
DT	01-MAR-2002 (Rel. 41, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Anthrax toxin receptor precursor (Tumor endothelial marker 8).			
GN	ATR OR TEM8.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RX	MEDLINE:20407466; PubMed:10947988;			
RA	St Croix B., Rago C., Velculescu V., Traverso G., Romans K.E.,			
RA	Montgomery E., Lai A., Riggins G.J., Lengauer C., Vogelstein B.,			
RA	Kinzler K.W.;			
RT	"Genes expressed in human tumor endothelium.";			
RL	Science 289:1197-1202(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A. (ISOFORM 2).			
RX	PubMed:11700362;			
RA	Bradley K.A., Mogridge J., Mourez M., Collier R.J., Young J.A.T.;			
RT	"Identification of the cellular receptor for anthrax toxin.";			
RL	Nature 414:225-229(2001).			
RN	[3]			
RP	SEQUENCE OF 184-564 FROM N.A. (ISOFORM 1), AND SEQUENCE FROM N.A. (ISOFORM 3).			
RA	Kawabata A., Hikiuji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,			
RA	Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,			
RA	Tanaka T., Nakamura Y., Isogai T., Sugano S.;			
RT	"NEDO human cDNA sequencing project.";			
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A. (ISOFORM 4).			
RX	TISSUE=Kidney;			
RA	Strausberg R.;			
RT	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.			
RL	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.			
CC	-!- FUNCTION: Cellular role is not yet known. (PA) of Bacillus anthracis.			
CC	-!- SUBUNIT: Binds to the protective antigen. (PA) of Bacillus anthracis.			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).			
CC	-!- ALTERNATIVE PRODUCTS: 4 isoforms; 1 (shown here), 2, 3 and 4; seem to be produced by alternative splicing.			
CC	-!- TISSUE SPECIFICITY: Highly expressed in tumor endothelial cells but not in normal endothelial cells.			
CC	-!- DOMAIN: Binding to PA seems to be effected through the VWA domain.			
CC	-!- SIMILARITY: BELONGS TO THE ATR FAMILY.			
CC	-!- SIMILARITY: CONTAINS 1 VWA DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way			

34	97	6.3	285	1	ITAI_CHICK	Q90615 gallus gall
35	97	6.3	3137	1	CA36_CHICK	P15989 gallus gall
36	96	6.2	639	1	CA1C_RABIT	Q28902 oryctolagus
37	95.5	6.2	1179	1	ITAE_HUMAN	P38570 homo sapien
38	95	6.2	1167	1	ITAE_MOUSE	Q06077 mus musculus
39	94.5	6.2	1151	1	ITAI_HUMAN	P56199 homo sapien
40	92.5	6.0	341	1	Y173_BORBU	O51195 borrelia bu
41	92.5	6.0	550	1	COCH_HUMAN	O43405 homo sapien
42	92.5	6.0	642	1	DNAK_NEIMA	Q91VQ9 neisseria m
43	91.5	6.0	1180	1	ITAI_RAT	P18614 rattus norv
44	91	5.9	1167	1	ITAG_HUMAN	O75578 homo sapien
45	90	5.9	1170	1	ITA2_BOVIN	P53710 bos taurus

## ALIGNMENTS

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AF279145; AAK52094.1; -;  
DR EMBL; AK025429; BAB15128.1; ALT\_INIT.  
DR EMBL; AK001463; BAA91707.1; ALT\_FRAME.  
DR EMBL; BC012074; AAH12074.1; -;  
DR MIM; 606410; -;

DR InterPro: IPR002035; VWFA.  
DR SMART; SM00327; VWA; 1.  
DR PROSITE; PS50234; VWFA; 1.  
KW Receptor; Transmembrane; Glycoprotein; Signal; Alternative splicing.

FT SIGNAL 1 32 POTENTIAL.  
FT CHAIN 33 564 ANTRAX TOXIN RECEPTOR.  
FT DOMAIN 33 321 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 322 342 POTENTIAL.  
FT DOMAIN 343 564 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 44 215 VWFA.  
FT DOMAIN 360 368 ASP/GLU-RICH (HIGHLY ACIDIC).  
FT DOMAIN 506 564 PRO-RICH.  
FT CARBOHYD 166 186 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 365 368 EDDD -> NKIK (IN ISOFORM 2).  
FT VARSPLIC 369 564 MISSING (IN ISOFORM 2).  
FT VARSPLIC 268 297 NEKPSFVEDTYLLCPAPILKEVGMKAALQV -> SKSLQSP  
FT FT MISSING (IN ISOFORM 3).  
FT FT WYSTSGFEGSHPCLPAPRPT (IN ISOFORM 3).  
FT FT MISSING (IN ISOFORM 3).  
FT FT DGSILAIALILFLL -> LHKTAGSPTTAACME (IN  
FT FT ISOFORM 4).  
FT VARSPLIC 334 564 MISSING (IN ISOFORM 4).  
FT SEQUENCE 564 AA; 62789 MW; B118A00AD5DF2233 CRC64;

Query Match 100.0%; Score 1536; DB 1; Length 564;  
Best Local Similarity 100.0%; Pred. No. 6.le-120;  
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GQGRREDGGPACGGFDLYFLDKSGSVLHHWNEIYFVEQLAKHFIISPOLRMSFIVFS 60  
Db 27 GQGRREDGGPACGGFDLYFLDKSGSVLHHWNEIYFVEQLAKHFIISPOLRMSFIVFS 86  
Qy 61 TRGTTMLKLTEDRQIRQLEELQKVLPGGDTYMHGFERASEQIYYENRGYRTASVII 120  
Db 87 TRGTTMLKLTEDRQIRQLEELQKVLPGGDTYMHGFERASEQIYYENRGYRTASVII 146  
Qy 121 ALTDGELHEDLFFYSERANRSDIGAIVCVGVKDFNETOLARIADSKDHVFPVNDGQF 180  
Db 147 ALTDGELHEDLFFYSERANRSDIGAIVCVGVKDFNETOLARIADSKDHVFPVNDGQF 206  
Qy 181 ALQGIHSILKSCIEILAAEPSITCAGESPQVVVRGNGFRHARNVDRVLCGSKINDSVT 240  
Db 207 ALQGIHSILKSCIEILAAEPSITCAGESPQVVVRGNGFRHARNVDRVLCGSKINDSVT 266  
Qy 241 LNEKPSFVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFSSVLIITTHCSGDS 295  
Db 267 LNEKPSFVEDTYLLCPAPILKEVGMKAALQVSMNDGLSFSSVLIITTHCSGDS 321

## RESULT 2

ID	ATR_MOUSE	STANDARD;	PRT;	562 AA.
AC	Q9CZ52;			
DT	01-MAR-2002 (Rel. 41, Created)			
DT	01-MAR-2002 (Rel. 41, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Antrax toxin receptor precursor (tumor endothelial marker 8).			
GN	ATR OR TM8.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
OX	NCBI_TaxID=10090;			

RN SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=21443268; PubMed=11559528;  
RA Carson-Walter E.B., Watkins D.N., Nanda A., Vogelstein B.,  
RA Kinzler K.W., St Croix B.;  
RT "Cell surface tumor endothelial markers are conserved in mice and  
RL humans.";  
RL Cancer Res. 61:6649-6655(2001).  
RN [2]  
RP SEQUENCE OF 88-562 FROM N.A. (ISOFORM 2).  
RC STRAIN=C57BL/6J; TISSUE=Embryo;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
CC -!- FUNCTION: Cellular role is not yet known.  
CC -!- SUBUNIT: Binds to the protective antigen (PA) of Bacillus  
CC anthracis (By similarity).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).  
CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; seem to  
CC be produced by alternative splicing.  
CC -!- DOMAIN: Binding to PA seems to be effected through the VWA domain  
CC (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE ATR FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF378762; AAL11999.1; -;  
DR EMBL; AK013005; BAB28591.1; ALT\_INIT.  
DR MGD; MGI:1919432; Atr.  
DR InterPro: IPR002035; VWFA.  
DR PROSITE; PS50234; VWFA; 1.  
KW Receptor; Transmembrane; Glycoprotein; Signal; Alternative splicing.  
FT SIGNAL 1 30 POTENTIAL.  
FT CHAIN 31 562 ANTRAX TOXIN RECEPTOR.  
FT DOMAIN 31 319 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 320 340 POTENTIAL.  
FT DOMAIN 341 562 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 42 213 VWFA.  
FT DOMAIN 358 366 ASP/GLU-RICH (HIGHLY ACIDIC).  
FT DOMAIN 501 562 PRO-RICH.  
FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARSPLIC 477 562 GRCINFTRVNSQAPKYPKPLNNYTHPSPPAPITPPPPAP  
FT FT RCPSPAPAPTTPSPPTLPPPPQAPPPNRRAPPPSRPP  
FT FT RPSV -> RFRGRLTICLSKGVHGRHDKGPETPLKQA  
FT FT WMPSSFLERAFQ (IN ISOFORM 2).  
FT SEQUENCE 562 AA; 62308 MW; 6AC92049B4B4F7C CRC64;



"The leukocyte integrin alpha D beta 2 binds VCAM-1: evidence for a binding interface between I domain and VCAM-1.";  
J. Immunol. 163:1984-1990(1999).  
-!- FUNCTION: INTEGRIN ALPHA-D/BETA-2 IS A RECEPTOR FOR ICAM3 AND VCAM1. MAY PLAY A ROLE IN THE ATHEROSCLEROTIC PROCESS SUCH AS CLEARING LIPOPROTEINS FROM PLAQUES AND IN PHAGOCYTOSIS OF BLOOD-BORNE PATHOGENS, PARTICULATE MATTER, AND SENESCENT ERYTHROCYTES FROM THE BLOOD.  
-!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-D ASSOCIATES WITH BETA-2.  
-!- SUBCELLULAR LOCATION: Type I membrane protein.  
-!- TISSUE SPECIFICITY: EXPRESSED MODERATELY ON MYELOMONOCYTIC CELL LINES AND SUBSETS OF PERIPHERAL BLOOD LEUKOCYTES AND STRONGLY ON TISSUE-SPECIALIZED CELLS, INCLUDING MACROPHAGES FOAM CELLS WITHIN ATHEROSCLEROTIC PLAQUES, AND ON SPLENIC RED PULP MACROPHAGES.  
-!- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
-!- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
-!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
-!- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; U37028; AAB38547.1; -  
DR EMBL; U40274; AAB60634.1; -  
DR EMBL; U40275; AAB60635.1; -  
DR EMBL; U40276; AAB60636.1; -  
DR EMBL; U40277; AAB60637.1; -  
DR EMBL; U40279; AAB60638.1; -  
DR EMBL; U40278; AAB60638.1; JOINED.  
DR EMBL; AF187881; AAF62875.1; -  
DR HSSP; P11215; 1A8X.  
DR MIM; 602453; -  
DR InterPro; IP000413; Integrin\_alpha.  
DR InterPro; IP002035; VWFA.  
DR Pfam; PR01639; FG-GAP; 5.  
DR Pfam; PF00357; Integrin\_A; 1.  
DR Pfam; PF00092; vwa; 1.  
DR PRINTS; PR01185; INTEGRINA.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00191; Int.alpha; 4.  
DR SMART; SM00327; VWA; 1.  
DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
DR PROSITE; PS50234; VWFA; 1.  
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
KW Signal; Repeat; Calcium;  
KW Magnesium.  
FT SIGNAL 1 17 POTENTIAL.  
FT CHAIN 18 1162 INTEGRIN ALPHA-D.  
FT DOMAIN 18 1100 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1101 1124 POTENTIAL.  
FT DOMAIN 1125 1162 CYTOPLASMIC (POTENTIAL).  
FT REPEAT 32 85 FG-GAP 1.  
FT REPEAT 2 2 FG-GAP 2.  
FT REPEAT 2 2 VWFA.  
FT DOMAIN 150 332 FG-GAP 3.  
FT REPEAT 350 400 FG-GAP 4.  
FT REPEAT 401 452 FG-GAP 5.  
FT REPEAT 454 516 FG-GAP 6.  
FT REPEAT 518 576 FG-GAP 7.  
FT REPEAT 581 633 POTENTIAL.  
FT CA\_BIND 465 473 POTENTIAL.  
FT CA\_BIND 530 538 POTENTIAL.  
FT CA\_BIND 593 601 POTENTIAL.  
FT SITE 1127 1131 GFGR MOTIF.  
FT SITE 67 74 BY SIMILARITY.  
FT DISULFID 106 124 BY SIMILARITY.  
FT DISULFID 655 710 BY SIMILARITY.

FT DISULFID 769 775 BY SIMILARITY.  
FT DISULFID 846 861 BY SIMILARITY.  
FT DISULFID 994 1018 BY SIMILARITY.  
FT DISULFID 1023 1028 BY SIMILARITY.  
FT CARBOHYD 59 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 87 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 99 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 391 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 691 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 733 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 873 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 957 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1046 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CONFLICT 500 MISSING (IN REF. 2).  
FT CONFLICT 515 GHPW -> ATP (IN REF. 2).  
FT CONFLICT 825 L -> V (IN REF. 2).  
FT CONFLICT 984 V -> A (IN REF. 2).  
SQ SEQUENCE 1162 AA; 126885 MW; F296A1A35455D77D CRC64;  
  
Query Match 10.2%; Score 156; DB 1; Length 1162;  
Best Local Similarity 24.3%; Pred. No. 4.4e-05;  
Matches 67; Conservative 50; Mismatches 119; Indels 40; Gaps 11;  
  
QY 8 DGGPAC-YGGFDLYFLDKSGSV-LHHWNEIYFYVEQLAKHFKISPOLRMSFIVFSTRGTT 65  
DB 139 DATPECPHGMEDIVFLIDGSGSIDQNDQNMKGFOVQAVMGQFEGDTLFLALMOYS-----N 194  
QY 66 LMKLTREDQIQGLELQKVLPP-----GGDTYMHGFERASEQIYYENRQGYRTA-SVII 120  
DB 195 LUKIHFTTQFRTSPSQQLSDIVQLKGLTFTATGILIVTQLFHKNKARKSAKKILI 254  
QY 121 ALTGDGELHEDLFYISREANRSRDLAGIYVCVGVKDFNETQLAR-----TADSKDHVF 173  
DB 255 VITDGQKYKDPLEYSYD-VIPOAEKAGIIRYAIGVGHAFCQPTARQELNITISSAPPQDHYF 313  
QY 174 PYNDGFOALQGIHSLKKSCTEILAAEPSTICAGSFQVWVRGNGFHRARNVDVLC-- 231  
DB 314 KV-DNFAALGSIQKQLOEK-----IYAVEGTSQRASSFQHEMSQEGFSTALTMGDLFGA 368  
QY 232 --SFKINDS-----VTLNKPFSEVDYTL 253  
DB 369 VGSFWSGGAFLYPPNMSPTFINNSQENVDMDRDSYL 404  
  
RESULT 5  
ID ITAM\_HUMAN STANDARD; PRT; 1152 AA.  
AC P11215;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Integrin alpha-M precursor (Cell surface glycoprotein MAC-1 alpha subunit) (CR-3 alpha chain) (CD11b) (Leukocyte adhesion receptor MO1)  
DE (Neutrophil adherence receptor).  
GN ITGAM OR CR3A OR CD11B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RP MEDLINE=8815033; PubMed=2457584;  
RA Corbi A.L., Kishimoto T.K., Miller L.J., Springer T.A.;  
RT "The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD11b) alpha subunit. Cloning, primary structure, and relation to the integrins, von Willebrand factor and factor B.";  
RL J. Biol. Chem. 263:12403-12411(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RP MEDLINE=88190151; PubMed=2833753;  
RA Arnaout M.A., Remold-O'Donnell E., Pierce M.W., Harris P., Tenen D.G.;  
RT "Molecular cloning of the alpha subunit of human and guinea pig



leukocyte adhesion glycoprotein Mol: chromosomal localization and  
homology to the alpha subunits of integrins.";  
Proc. Natl. Acad. Sci. U.S.A. 85:2776-2780(1988).  
[3]  
SEQUENCE FROM N.A.  
RP MEDLINE=98257215; PubMed=2454931;  
RX Arnaout M.A., Gupta S.K., Pierce M.W., Tenen D.G.;  
RA "Amino acid sequence of the alpha subunit of human leukocyte adhesion  
RT receptor Mol (complement receptor type 3).";  
RL J. Cell Biol. 106:2153-2158(1988).  
[4]  
SEQUENCE FROM N.A.  
RP MEDLINE=93123748; PubMed=8419480;  
RX Fleming J.C., Pahl H.L., Gonzalez D.A., Smith T.F., Tenen D.G.;  
RA "Structural analysis of the CD11b gene and phylogenetic analysis of  
RT the alpha-integrin gene family demonstrate remarkable conservation of  
RL genomic organization and suggest early diversification during  
evolution.";  
RJ. Immunol. 150:480-490(1993).  
[5]  
SEQUENCE OF 1-9 FROM N.A.  
RP MEDLINE=9098893; PubMed=2563162;  
RX Hickstein D.D., Hickey M.J., Ozols J., Baker D.M., Back A.L.,  
RA Roth G.J.;  
RT "cDNA sequence for the alpha M subunit of the human neutrophil  
adherence receptor indicates homology to integrin alpha subunits.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:257-261(1989).  
[6]  
SEQUENCE OF 1-9 FROM N.A.  
RP MEDLINE=92073318; PubMed=1683702;  
RX Shelley C.S., Arnaout M.A.;  
RA "The promoter of the CD11b gene directs myeloid-specific and  
RT developmentally regulated expression.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:10525-10529(1991).  
[7]  
SEQUENCE OF 1-9 FROM N.A.  
RP TISSUE=Blood;  
RX MEDLINE=92144986; PubMed=1346576;  
RA Pahl H.L., Rosmarin A.G., Tenen D.G.;  
RT "Characterization of the myeloid-specific CD11b promoter.";  
RL Blood 79:865-870(1992).  
[8]  
SEQUENCE OF 17-31.  
RP MEDLINE=87076671; PubMed=3539202;  
RX Pierce M.W., Remold-O'Donnell E., Todd R.F., III, Arnaout M.A.;  
RA "N-terminal sequence of human leukocyte glycoprotein Mol:  
RT conservation across species and homology to platelet IIb/IIIa.";  
RL Blochim. Biophys. Acta 874:368-371(1986).  
[9]  
X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 148-331.  
RP MEDLINE=95171458; PubMed=7867070;  
RX Lee J.O., Rieu P., Arnaout M.A., Liddington R.;  
RA "Crystal structure of the A domain from the alpha subunit of integrin  
RT C3 (CD11b/CD18).";  
RL Cell 80:631-638(1995).  
[10]  
X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 148-334.  
RP MEDLINE=96363671; PubMed=8747460;  
RX Lee J.O., Bankston L.A., Arnaout M.A., Liddington R.C.;  
RA "Two conformations of the integrin A-domain (I-domain): a pathway for  
RT activation?";  
RL Structure 3:1333-1340(1995).  
[11]  
X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 148-337.  
RP MEDLINE=98362595; PubMed=9687375;  
RX Baldwin E.T., Sarver R.W., Bryant G.L. Jr., Curry K.A.,  
RA Fairbanks M.B., Finzel B.C., Garlick R.L., Heinrichson R.L.,  
RA Horton N.C., Kelley L.L., Mildner A.M., Moon J.B., Mott J.E.,  
RA Mutchler V.T., Tomich C.S., Watenpaugh K.D., Wiley V.H.;  
RT "Cation binding to the integrin CD11b I domain and activation model  
assessment.";  
RL Structure 6:923-935(1998).  
[12]  
3D-STRUCTURE MODELING OF 17-616.  
RP MEDLINE=98226734; PubMed=9560195;  
RX Oxvig C., Springer T.A.;  
RT "Experimental support for a beta-propeller domain in integrin alpha-  
subunits and a calcium binding site on its lower surface.";  
Proc. Natl. Acad. Sci. U.S.A. 95:4870-4875(1998).  
[13]  
FUNCTION: INTEGRIN ALPHA-M/BETA-2 IS IMPLICATED IN VARIOUS  
ADHESIVE INTERACTIONS OF MONOCYTES, MACROPHAGES AND GRANULOCYTES  
AS WELL AS IN MEDIATING THE UPTAKE OF COMPLEMENT-COATED PARTICLES.  
IT IS IDENTICAL WITH CR-3. THE RECEPTOR FOR THE IC3B FRAGMENT OF  
THE THIRD COMPLEMENT COMPONENT. IT PROBABLY RECOGNIZES THE R-G-D  
PEPTIDE IN C3B. INTEGRIN ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR  
OF FIBRINOGEN, FACTOR X AND ICAM1. IT RECOGNIZES P1 AND P2 PEPTIDES  
ASSOCIATES WITH BETA-2.  
[14]  
SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-M  
[15]  
SUBCELLULAR LOCATION: Type I membrane protein.  
[16]  
TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND  
GRANULOCYTES.  
[17]  
DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VFHA DOMAIN. INTEGRINS  
WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
[18]  
SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
[19]  
SIMILARITY: CONTAINS 1 VFHA DOMAIN.  
[20]  
SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
[21]  
DATABASE: NAME-PROW; NOTE-CD guide CD11b entry;  
WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd11b.htm".  
[22]  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
[23]  
EMBL; J03925; AAA59544.1; -;  
EMBL; M18044; AAA59491.1; -;  
EMBL; J04145; AAA59903.1; -;  
EMBL; S52227; AAB24821.1; -;  
EMBL; S52152; AAB24821.1; JOINED.  
EMBL; S52153; AAB24821.1; JOINED.  
EMBL; S52154; AAB24821.1; JOINED.  
EMBL; S52155; AAB24821.1; JOINED.  
EMBL; S52157; AAB24821.1; JOINED.  
EMBL; S52159; AAB24821.1; JOINED.  
EMBL; S52161; AAB24821.1; JOINED.  
EMBL; S52164; AAB24821.1; JOINED.  
EMBL; S52165; AAB24821.1; JOINED.  
EMBL; S52167; AAB24821.1; JOINED.  
EMBL; S52169; AAB24821.1; JOINED.  
EMBL; S52170; AAB24821.1; JOINED.  
EMBL; S52173; AAB24821.1; JOINED.  
EMBL; S52174; AAB24821.1; JOINED.  
EMBL; S52180; AAB24821.1; JOINED.  
EMBL; S52181; AAB24821.1; JOINED.  
EMBL; S52184; AAB24821.1; JOINED.  
EMBL; S52189; AAB24821.1; JOINED.  
EMBL; S52191; AAB24821.1; JOINED.  
EMBL; S52192; AAB24821.1; JOINED.  
EMBL; S52203; AAB24821.1; JOINED.  
EMBL; S52212; AAB24821.1; JOINED.  
EMBL; S52213; AAB24821.1; JOINED.  
EMBL; S52216; AAB24821.1; JOINED.  
EMBL; S52219; AAB24821.1; JOINED.  
EMBL; S52220; AAB24821.1; JOINED.  
EMBL; S52221; AAB24821.1; JOINED.  
EMBL; S52222; AAB24821.1; JOINED.  
EMBL; S52226; AAB24821.1; JOINED.  
EMBL; M76724; AAA58410.1; -;  
EMBL; M84477; AAA51960.1; -;  
PIR; A31108; RWHUIB.  
PIR; A26091; A26091.  
PDB; 1ABX; 17-JUN-98.

DR PDB; 1BHO; 18-NOV-98.  
DR PDB; 1BHQ; 18-NOV-98.  
DR PDB; 1IDN; 25-NOV-98.  
DR PDB; 1IDO; 01-AUG-96.  
DR PDB; 1JLM; 11-JAN-97.  
DR CarBbank; CCSD:33581; -.  
DR CarBbank; CCSD:33582; -.  
DR CarBbank; CCSD:33584; -.  
DR MIM; 120980; -.  
DR InterPro; IPR000413; Integrin\_alpha.  
DR InterPro; IPR002035; VWFA.  
DR Pfam; PF01839; FG-GAP; 5.  
DR Pfam; PF00357; Integrin\_A; 1.  
DR Pfam; PF00092; vwa; 1.  
DR PRINTS; PR01185; INTEGRINA.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00191; Int\_alpha; 4.  
DR SMART; SM00327; VWA; 1.  
DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
DR PROSITE; PS0234; VWFA; 1.  
DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
KW Signal; 3D-structure; Repeat; Magnesium; Calcium.  
FT SIGNAL 1 16  
FT CHAIN 17 1152  
FT DOMAIN 17 1104  
FT TRANSMEM 1105 1128  
FT DOMAIN 1129 1152  
FT REPEAT 31 84  
Query Match 9.5%; Score 145.5; DB 1; Length 1152;  
Best Local Similarity 26.2%; Pred. No. 0.00032;  
Matches 60; Conservative 45; Mismatches 81; Indels 43; Gaps 12;  
QY 18 DLYFILDKGSVL-HWNNEIYYEQLAHKFISPOLRMSFVSTGTTLMKLFED--- 72  
Db 150 DIAFLDGSIIIPHFRKKEFVST-----VMEQLKSKTFLS-----LMQSEEFRIH 199  
QY 73 -----RQIRQGLEELQKVLPGDGYMHGFERASEQIYYENRQGYRTA-SVIAITL 123  
Db 200 FTFKEQNPNRSLVKPIQTQLL--GRTHPTGIRKVVRELENTTGARKNAFKILVIT 257  
QY 124 DGEHLEDLFYSE--REANRSRLGAIVYGVKDFNETQLAR-----IADS--KDHVFP 174  
Db 258 DGEKFGDPLGVDEIVPEADRE--GVIRYVIGVDAPFRSEKSRQELNTIASKPPRDHVQ 314  
QY 175 VNDGFOALOGIIHSILKSCIEILAEPTICAGSFQVVGNGFRHA 223  
Db 315 VNN-FEALKTIQNLREK----IFAIEGTQTGSSSFEHEMSQEGFSAA 358  
RESULT 6  
ITAX\_HUMAN STANDARD; PRT; 1163 AA.  
AC F20702;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DE 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Integrin alpha-x precursor (Leukocyte adhesion glycoprotein p150,95  
DE alpha chain) (Leukocyte adhesion receptor p150,95) (CD11c) (Leu M5).  
GN ITGAX OR CD11C.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88166645; PubMed=3327687;  
RA Corbi A.L., Miller L.J., O'Connor K., Larson R.S., Springer T.A.;  
RT "cDNA cloning and complete primary structure of the alpha subunit of  
RT a leukocyte adhesion glycoprotein, p150,95.";  
RL EMBO J. 6:4023-4028(1987).  
RN [2]  
RP SEQUENCE FROM N.A.

RX MEDLINE=90153906; PubMed=2303426;  
RA Corbi A.L., Garcia-Aguilar J., Springer T.A.;  
RT "Genomic structure of an integrin alpha subunit, the leukocyte  
RT p150,95 molecule.";  
RL J. Biol. Chem. 265:2782-2788(1990).  
RN [3]  
RP ERRATUM.  
RA Corbi A.L., Garcia-Aguilar J., Springer T.A.;  
RL J. Biol. Chem. 265:12750-12751(1990).  
RN [4]  
RP SEQUENCE OF 20-43.  
RX MEDLINE=87167596; PubMed=3549901;  
RA Miller L.J., Wiebe M., Springer T.A.;  
RT "Purification and alpha subunit N-terminal sequences of human Mac-1  
RT and p150,95 leukocyte adhesion proteins.";  
RL J. Immunol. 138:2381-2383(1987).  
CC -!- FUNCTION: INTEGRIN ALPHA-X/BETA-2 IS A RECEPTOR FOR FIBRINOGEN. IT  
CC RECOGNIZES THE SEQUENCE G-P-R IN FIBRINOGEN. IT MEDIATES CELL-CELL  
CC INTERACTION DURING INFLAMMATORY RESPONSES. IT IS ESPECIALLY  
CC IMPORTANT IN MONOCYTE ADHESION AND CHEMOTAXIS.  
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-X  
CC ASSOCIATES WITH BETA-2.  
CC -!- TISSUE SPECIFICITY: Type I membrane protein.  
CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND  
CC GRANULOCYTES.  
CC -!- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS  
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
CC -!- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
CC -!- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
CC -!- DATABASE: NAME=PROW; NOTE=CD guide CD11c entry;  
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd11c.htm".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; M81695; AAA59180.1; -.  
DR EMBL; Y00093; CAA68283.1; -.  
DR EMBL; M29165; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; M29487; AAA51620.1; ALT\_SEQ.  
DR EMBL; M29482; AAA51620.1; JOINED.  
DR EMBL; M29483; AAA51620.1; JOINED.  
DR EMBL; M29484; AAA51620.1; JOINED.  
DR EMBL; M29485; AAA51620.1; JOINED.  
DR EMBL; M29486; AAA51620.1; JOINED.  
DR PIR; A36584; RWHUIC.  
DR HSSP; P11215; 1A8X.  
DR CarBbank; CCSD:33581; -.  
DR CarBbank; CCSD:33582; -.  
DR CarBbank; CCSD:33584; -.  
DR MIM; 151510; -.  
DR InterPro; IPR000413; Integrin\_alpha.  
DR InterPro; IPR002035; VWFA.  
DR Pfam; PF01839; FG-GAP; 5.  
DR Pfam; PF00357; Integrin\_A; 1.  
DR Pfam; PF00092; vwa; 1.  
DR PRINTS; PR01185; INTEGRINA.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00191; Int\_alpha; 4.  
DR SMART; SM00327; VWA; 1.  
DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
DR PROSITE; PS0234; VWFA; 1.  
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
KW Signal; Magnesium; Calcium; Repeat.  
FT SIGNAL 1 19  
FT CHAIN 20 1163  
FT DOMAIN 20 1107  
FT TRANSMEM 1108 1128  
FT POTENTIAL.

```
FT DOMAIN 1129 1163 CYTOPLASMIC (POTENTIAL).
FT REPEAT 34 87 FG-GAP 1.
FT REPEAT ? 2 FG-GAP 2.
FT DOMAIN 165 351 VWFA.
FT REPEAT ? 453 FG-GAP 3.
FT REPEAT 402 453 FG-GAP 4.
FT REPEAT 455 517 FG-GAP 5.
FT REPEAT 518 576 FG-GAP 6.
FT REPEAT 581 633 FG-GAP 7.
FT CA_BIND 466 474 POTENTIAL.
FT CA_BIND 530 538 POTENTIAL.
FT CA_BIND 593 601 POTENTIAL.
FT SITE 1131 1135 GFFKR MOTIF.
FT DISULFID 69 76 BY SIMILARITY.
FT DISULFID 108 126 BY SIMILARITY.
FT DISULFID 655 712 BY SIMILARITY.
FT DISULFID 771 777 BY SIMILARITY.
FT DISULFID 848 863 BY SIMILARITY.
FT DISULFID 998 1022 BY SIMILARITY.
FT DISULFID 1027 1032 BY SIMILARITY.
FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 697 697 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 735 735 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 939 939 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1050 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 490 490 G -> A (IN REF. 2).
FT CONFLICT 756 756 L -> D (IN REF. 2).
SQ SEQUENCE 1163 AA; 127885 MW; 6C4E19CC3F62A473 CRC64;

Query Match 9.38; Score 143.5; DB 1; Length 1163;
Best Local Similarity 24.88; Pred. No. 0.00048;
Matches 55; Conservative 43; Mismatches 89; Indels 35; Gaps 10;

Qy 18 DLYFLDKSGSV-LHHWNEIYFYFQQLAHKFTISPOLRMSETVFSPRGTTLMKLTEDRE-- 74
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 151 DIVFLDGGSGSISSENFATMNFRAVISQFQSTQSLQFQSKFQTHFTFEFRFTS 210
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 75 ---QIROGLEELQKVLPGDQTYMHGFEFRASEQIYYENRQGYRTAS-VIIALTQGLHED 130
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 211 NPLSLASVHQLQ-----GFTYTATAIGNVVRHLFPHASYGARRDATKILIVTDGKKEG 265
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 131 LFFYSERANRDLGAIVYCVGV-----KDFNETQLARIAD--SKDHVFPVNDG 178
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 266 SLDYKD-VIPMDAAGIIRYAGVGLAFQNRNSWKELND-----IASKPSQEHIFKVED- 318
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 179 FOALQGHSHILKSKSCIEILAAEPSTICAGESFOVVVRGNFG 220
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 FDALKDIGNQLKEK-----IFAIEGTETTSSSSSFELEMAQEGF 356
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
CA1C_CHICK STANDARD; PRT; 3124 AA.
AC P13944; Q04509;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(XII) chain precursor (Fibrochimerin).
GN COL12A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WHITE LEIGHOR;
RX MEDLINE=92011862; PubMed=1918137;
RA Yamagata M., Yamada K.M., Yamada S.S., Shinomura T., Tanaka H.,
-----
Nishida Y., Obara M., Kimata K.;
"the complete primary structure of type XII collagen shows a chimeric
molecule with reiterated fibronectin type III motifs, von Willebrand
factor A motifs, a domain homologous to a noncollagenous region of
type IX collagen, and short collagenous domains with an Arg-Gly-Asp
site.";
J. Cell Biol. 115:209-221(1991).
[2]
SEQUENCE OF 2456-3124 FROM N.A., AND SEQUENCE OF 2772-2794 AND
2846-2873.
MEDLINE=90062079; PubMed=2584192;
Gordon M.K., Gerecke D.R., Dublet B., van der Rest M., Olsen B.R.;
"Type XII collagen. A large multidomain molecule with partial
homology to type IX collagen.";
J. Biol. Chem. 264:19772-19778(1989).
[3]
SEQUENCE OF 2960-3076 FROM N.A.
MEDLINE=87317590; PubMed=3476925;
Gordon M.K., Gerecke D.R., Olsen B.R.;
"Type XII collagen: distinct extracellular matrix component
discovered by cDNA cloning.";
Proc. Natl. Acad. Sci. U.S.A. 84:6040-6044(1987).
[4]
SEQUENCE OF 1-1283 FROM N.A. (SHORT FORM), AND ALTERNATIVE SPLICING.
TISSUE=Embryo;
MEDLINE=93042014; PubMed=1420368;
Trueb J., Trueb B.;
"The two splice variants of collagen XII share a common 5' end.";
Biochim. Biophys. Acta 1171:97-98(1992).
[5]
ALTERNATIVE SPLICING.
MEDLINE=95370352; PubMed=7642694;
Koch M., Bohmann B., Matthison M., Hagios C., Trueb B., Chiquet M.;
"Large and small splice variants of collagen XII: differential
expression and ligand binding.";
J. Cell Biol. 130:1005-1014(1995).
CC -!- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-
CONTAINING FIBRILS. THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE
SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE
LOCALIZED IN THE PERIFIBRILLAR MATRIX.
CC -!- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 kDa OF
NONTRIPLE-HELICAL SEQUENCES.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE FINAL TISSUE
FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OF EITHER THE LONGER
OR THE SHORTER ISOFORM OR ANY COMBINATION OF LONG AND SHORT
ISOFORM CHAINS. ONLY THE LONG VARIANT IS A PROTEOLYCAN. THE LARGE
ISOFORM HAS MORE RESTRICTED EXPRESSION IN EMBRYONIC TISSUE THAN
THE SMALL.
CC -!- TISSUE SPECIFICITY: TYPE XII COLLAGEN IS PRESENT IN TENDONS,
LIGAMENTS, PERICHONDRUM, AND PERIOSTEUM, ALL DENSE CONNECTIVE
TISSUES CONTAINING TYPE I COLLAGEN.
CC -!- DOMAIN: THIS SEQUENCE DEFINES FIVE DISTINCT DOMAINS, TWO TRIPLE-
HELICAL DOMAINS (COL1 AND COL2) AND THREE NONTRIPLE-HELICAL
DOMAINS (NC1, NC2, AND NC3).
CC -!- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT
EACH END.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
INTERRUPTED HELICES (FACIT) FAMILY.
CC -!- SIMILARITY: CONTAINS 4 VWFA DOMAINS.
CC -!- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
```



CC	-1- FUNCTION: INTEGRIN ALPHA-M/BETA-2 IS IMPLICATED IN VARIOUS
CC	ADHESIVE INTERACTIONS OF MONOCYTES, MACROPHAGES AND GRANULOCYTES
CC	AS WELL AS IN MEDIATING THE UPTAKE OF COMPLEMENT-COATED PARTICLES.
CC	IT IS IDENTICAL WITH CR-3, THE RECEPTOR FOR THE IC3B FRAGMENT OF
CC	THE THIRD COMPLEMENT COMPONENT. IT PROBABLY RECOGNIZES THE R-G-D
CC	PEPTIDE IN C3B. INTEGRIN ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR
CC	FIBRINOGEN FACTOR X AND ICAM1. IT RECOGNIZES PI AND P2 PEPTIDES
CC	OF FIBRINOGEN GAMMA CHAIN. ALPHA-M/BETA-2 PLAY A CRITICAL ROLE IN
CC	MAST CELL DEVELOPMENT AND IN IMMUNE COMPLEX-MEDIATED
CC	GLOMERULONEPHRITIS. MICE EXPRESSING A NULL MUTATION OF THE ALPHA-M
CC	SUBUNIT GENE DEMONSTRATE INCREASE IN NEUTROPHIL ACCUMULATION, IN
CC	RESPONSE TO A IMPAIRED DEGRANULATION AND PHAGOCYTOSIS, EVENTS THAT
CC	APPARENTLY ACCELERATE APOPTOSIS IN NEUTROPHILS. THESE MICE DEVELOP
CC	OBESITY.
CC	-1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN. ALPHA-M
CC	ASSOCIATES WITH BETA-2.
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.
CC	-1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND
CC	GRANULOCYTES.
CC	-1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC	WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC	-1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC	-1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC	-1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; X07640; CAA30479.1; -.
DR	EMBL; M14293; AAA39484.1; -.
DR	PIR; S00551; S00551.
DR	HSP; P11215; IABX.
DR	MGI; 96607; Itgam.
DR	InterPro; IPR000413; Integrin_alpha.
DR	InterPro; IPR02035; VWFA.
DR	Pfam; PF01839; FG-GAP; 5.
DR	Pfam; PF00357; Integrin_A; 1.
DR	Pfam; PF00092; vwa; 1.
DR	PRINTS; PR01185; INTEGRINA.
DR	PRINTS; PR00453; VWFADOMAIN.
DR	SMART; SM00191; Int_alpha; 4.
DR	SMART; SM00327; vwa; 1.
DR	PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR	PROSITE; PS00234; VWFA; 1.
KW	Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW	Signal; Calcium; Repeat.
FT	SIGNAL 1 16
FT	CHAIN 17 1153
FT	DOMAIN 17 1105
FT	TRANSMEM 1106 1129
FT	DOMAIN 1130 1153
FT	REPEAT 31 84
FT	REPEAT ? ?
FT	DOMAIN 164 350
FT	REPEAT 337 400
FT	REPEAT 401 452
FT	REPEAT 454 515
FT	REPEAT 517 575
FT	REPEAT 580 632
FT	CA_BIND 465 473
FT	POTENTIAL.
FT	CA_BIND 529 537
FT	POTENTIAL.
FT	CA_BIND 592 600
FT	POTENTIAL.
FT	SITE 1132 1136
FT	GFGR MOTIF.
FT	BY SIMILARITY.
FT	DISULFID 66 73
FT	BY SIMILARITY.
FT	DISULFID 105 123
FT	BY SIMILARITY.
FT	DISULFID 654 711
FT	BY SIMILARITY.
FT	DISULFID 770 776
FT	BY SIMILARITY.
FT	DISULFID 999 1023
FT	BY SIMILARITY.

CC CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
 CC CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE FINAL TISSUE  
 CC CC FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OF EITHER THE LONGER  
 CC CC OR THE SHORTER ISOFORM OR ANY COMBINATION OF LONG AND SHORT  
 CC CC ISOFORM CHAINS. ONLY THE LONG VARIANT IS A PROTEOGLYCAN.  
 CC CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TENDONS, PERICHONDRUM,  
 CC CC SKIN, CORNEA, SCLERA, BLOOD VESSELS, AND PERIOSTEUM.  
 CC CC -!- DEVELOPMENTAL STAGE: LONG FORM OF COLLAGEN XII, XIIA, IS THE  
 CC CC PREDOMINANT FORM AT EARLY STAGES (ED7 AND 11); AT LATER STAGES OF  
 CC CC DEVELOPMENT (ED15 AND 17) THE SHORT FORM, XIIB, BECOMES THE MAJOR  
 CC CC FORM. AS THE SHORT FORM BECOMES THE MAJOR PRODUCT, THE LONG SPLICE  
 CC CC VARIANT CONTINUES TO BE EXPRESSED IN SEVERAL TISSUES, EVEN AFTER  
 CC CC BIRTH.  
 CC CC -!- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT  
 CC CC EACH END (BY SIMILARITY).  
 CC CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY  
 CC CC SIMILARITY).  
 CC CC -!- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY).  
 CC CC -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH  
 CC CC INTERRUPTED HELICES (FACIT) FAMILY.  
 CC CC -!- SIMILARITY: CONTAINS 4 VWFA DOMAINS.  
 CC CC -!- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC CC -----  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC CC the European Bioinformatics Institute. There are no restrictions on its  
 CC CC use by non-profit institutions as long as its content is in no way  
 CC CC modified and this statement is not removed. Usage by and for commercial  
 CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC -----  
 CC CC EMBL; U25652; AAA99719.1; -  
 CC CC HSSP; P02751; 1TTF.  
 CC CC MGD; MGI:88448; Coll12a1.  
 CC CC InterPro: IPR000087; Collagen.  
 CC CC InterPro: IPR003961; FN.III.  
 CC CC InterPro: IPR003962; FnIII\_repeat.  
 CC CC InterPro: IPR003129; TSPN.  
 CC CC InterPro: IPR002035; VWFA.  
 CC CC Pfam; PF01391; Collagen; 4.  
 CC CC Pfam; PF00041; fn3; 18.  
 CC CC Pfam; PF02210; TSPN; 1.  
 CC CC Pfam; PF00092; vwa; 4.  
 CC CC PRINTS; PR00014; FNTYPEIII.  
 CC CC PRINTS; PR00453; VWFADOMAIN.  
 CC CC SMART; SM00060; FN3; 16.  
 CC CC SMART; SM00210; TSPN; 1.  
 CC CC SMART; SM00327; VWA; 4.  
 CC CC PROSITE; PS50234; VWFA; 4.  
 CC CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 CC CC Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.  
 CC CC SIGNAL 1 24  
 CC CC POTENTIAL.  
 CC CC CHAIN 25 3067  
 CC CC COLLAGEN ALPHA 1(XII) CHAIN.  
 CC CC FIBRONECTIN TYPE-III 1.  
 CC CC VWFA 1.  
 CC CC DOMAIN 25 114  
 CC CC DOMAIN 140 316  
 CC CC DOMAIN 332 423  
 CC CC FIBRONECTIN TYPE-III 2.  
 CC CC VWFA 2.  
 CC CC DOMAIN 634 725  
 CC CC FIBRONECTIN TYPE-III 3.  
 CC CC FIBRONECTIN TYPE-III 4.  
 CC CC FIBRONECTIN TYPE-III 5.  
 CC CC FIBRONECTIN TYPE-III 6.  
 CC CC FIBRONECTIN TYPE-III 7.  
 CC CC FIBRONECTIN TYPE-III 8.  
 CC CC VWFA 3.  
 CC CC DOMAIN 1203 1375  
 CC CC FIBRONECTIN TYPE-III 9.  
 CC CC FIBRONECTIN TYPE-III 10.  
 CC CC FIBRONECTIN TYPE-III 11.  
 CC CC FIBRONECTIN TYPE-III 12.  
 CC CC FIBRONECTIN TYPE-III 13.  
 CC CC FIBRONECTIN TYPE-III 14.  
 CC CC FIBRONECTIN TYPE-III 15.  
 CC CC FIBRONECTIN TYPE-III 16.  
 CC CC DOMAIN 1388 1474  
 CC CC DOMAIN 1480 1568  
 CC CC DOMAIN 1509 1622  
 CC CC DOMAIN 1659 1757  
 CC CC DOMAIN 1758 1848  
 CC CC DOMAIN 1849 1938  
 CC CC DOMAIN 1939 2029  
 CC CC DOMAIN 2030 2120

FT DOMAIN 2121 2208 FIBRONECTIN TYPE-III 17.  
 FT DOMAIN 2209 2297 FIBRONECTIN TYPE-III 18.  
 FT DOMAIN 2329 2501 VWFA 4.  
 FT DOMAIN 2456 2751 NONHELICAL REGION (NC3).  
 FT DOMAIN 2752 2899 TRIPLE-HELICAL REGION (COL2)  
 FT WITH 1 IMPERFECTION.  
 FT DOMAIN 2900 2942 NONHELICAL REGION (NC2).  
 FT DOMAIN 2943 3045 TRIPLE-HELICAL REGION (COL1)  
 FT WITH 2 IMPERFECTIONS.  
 FT DOMAIN 3046 3064 NONHELICAL REGION (NC1).  
 FT BINDING 802 802 TO CHONDROITIN SULFATE (POTENTIAL).  
 FT BINDING 893 893 TO CHONDROITIN SULFATE (POTENTIAL).  
 FT BINDING 985 985 TO CHONDROITIN SULFATE (POTENTIAL).  
 FT SITE 866 868 CELL ATTACHMENT SITE (POTENTIAL).  
 FT SITE 2784 2786 CELL ATTACHMENT SITE (POTENTIAL).  
 FT SITE 2896 2898 CELL ATTACHMENT SITE (POTENTIAL).  
 FT MOD\_RES 2945 2945 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 2948 2948 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 2951 2951 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 2960 2960 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 2966 2966 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 2969 2969 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 2972 2972 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 2984 2984 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 3001 3001 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 3004 3004 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 3015 3015 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 3024 3024 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 3027 3027 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 3030 3030 HYDROXYLATION (BY SIMILARITY).  
 FT DOMAIN 869 872 POLY-THR.  
 FT CARBOHYD 704 704 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1769 1769 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2212 2212 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2533 2533 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2684 2684 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPPLIC 25 1190 MISSING (IN SHORT ISOFORM).  
 SQ SEQUENCE 3067 AA; 334094 MW; 3FB5DEFB8A2CDB95 CRC64;  
 Query Match 9.0%; Score 139; DB 1; Length 3067;  
 Best Local Similarity 30.3%; Pred. No. 0.0039;  
 Matches 60; Conservative 37; Mismatches 75; Indels 26; Gaps 12;  
 QY 18 DLYFILDKGSV-LHHWNEIYFVEQLAKHF-TSP-QLRMSFVSTRGTMLKLFEDRE 74  
 Db 444 DIVFLVDGSGYSGTANFVKVRAFLVLAKEVLSFSPNRVQSLVQYSRDPHTTEFLKFN 503  
 QY 75 QIRQGLEELQKVL-----PGDITYMHEGFERASEQIYYENRQGYRT--ASVIALTDGEL 127  
 Db 504 -----VEDIIKAINTPYRGSGSTNTGKAMTYVREKIFVPNK-GSRNVPKVMILLTDGE- 556  
 QY 128 HEDLFYFSEARENRSDLGAIVYCVGVKDFNETQLARIAD--SKDHVPVNDGFOALQGI 185  
 Db 557 SSDAF---RDPAILKNSDVEIFAVGVKDAVRSELEAISPAPAEHTVFTVED-FDAFORI 612  
 QY 186 IHSILKKSCIEI---LAA 200  
 Db 613 SFVLTSQICLRIRIEQLAA 630  
 RESULT 10  
 CO2\_MOUSE  
 ID CO2\_MOUSE STANDARD; PRT; 760 AA.  
 AC P21180;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Complement C2 precursor (EC 3.4.21.43) (C3/C5 convertase).  
 GN C2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCHI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91035430; PubMed=2229060;  
 RA Ishikawa N., Nonaka M., Wetzel R.A., Colten H.R.;  
 RT "Murine complement C2 and factor B genomic and cDNA cloning reveals  
 RT different mechanisms for multiple transcripts of C2 and B.";  
 RL J. Biol. Chem. 265:19040-19046(1990).  
 CC -!- FUNCTION: COMPONENT C2 WHICH IS PART OF THE CLASSICAL PATHWAY OF  
 CC THE COMPLEMENT SYSTEM IS CLEAVED BY ACTIVATED FACTOR C1 INTO TWO  
 CC FRAGMENTS: C2B AND C2A. C2A, A SERINE PROTEASE, THEN COMBINES WITH  
 CC COMPLEMENT FACTOR 4B TO GENERATE THE C3 OR C5 CONVERTASE.  
 CC -!- CATALYTIC ACTIVITY: CLEAVES C3 IN THE ALPHA-CHAIN TO YIELD C3A AND  
 CC C3B. CLEAVES C5 IN THE ALPHA-CHAIN TO YIELD C5A AND C5B. BOTH  
 CC CLEAVAGES TAKE PLACE AT THE C-TERMINAL OF AN ARGININE RESIDUE.  
 CC -!- MISCELLANEOUS: C2 IS A MAJOR HISTOCOMPATIBILITY COMPLEX CLASS-III  
 CC PROTEIN.  
 CC -!- SIMILARITY: WITH COMPLEMENT FACTOR B.  
 CC -!- SIMILARITY: CONTAINS 3 SUSHI (SCR) DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 1 VWFA DOMAIN  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M60579; AAA37380.1; -  
 CC EMBL; M60563; AAA37380.1; JOINED.  
 CC EMBL; M60564; AAA37380.1; JOINED.  
 CC EMBL; M60565; AAA37380.1; JOINED.  
 CC EMBL; M60566; AAA37380.1; JOINED.  
 CC EMBL; M60567; AAA37380.1; JOINED.  
 CC EMBL; M60568; AAA37380.1; JOINED.  
 CC EMBL; M60569; AAA37380.1; JOINED.  
 CC EMBL; M60570; AAA37380.1; JOINED.  
 CC EMBL; M60571; AAA37380.1; JOINED.  
 CC EMBL; M60572; AAA37380.1; JOINED.  
 CC EMBL; M60573; AAA37380.1; JOINED.  
 CC EMBL; M60574; AAA37380.1; JOINED.  
 CC EMBL; M60575; AAA37380.1; JOINED.  
 CC EMBL; M60605; AAA37380.1; JOINED.  
 CC EMBL; M60576; AAA37380.1; JOINED.  
 CC EMBL; M60577; AAA37380.1; JOINED.  
 CC EMBL; M60578; AAA37380.1; JOINED.  
 CC EMBL; M60579; AAA37381.1; -  
 CC EMBL; M60563; AAA37381.1; JOINED.  
 CC EMBL; M60564; AAA37381.1; JOINED.  
 CC EMBL; M60565; AAA37381.1; JOINED.  
 CC EMBL; M60566; AAA37381.1; JOINED.  
 CC EMBL; M60567; AAA37381.1; JOINED.  
 CC EMBL; M60568; AAA37381.1; JOINED.  
 CC EMBL; M60569; AAA37381.1; JOINED.  
 CC EMBL; M60570; AAA37381.1; JOINED.  
 CC EMBL; M60571; AAA37381.1; JOINED.  
 CC EMBL; M60572; AAA37381.1; JOINED.  
 CC EMBL; M60573; AAA37381.1; JOINED.  
 CC EMBL; M60574; AAA37381.1; JOINED.  
 CC EMBL; M60575; AAA37381.1; JOINED.  
 CC EMBL; M60605; AAA37381.1; JOINED.  
 CC EMBL; M60576; AAA37381.1; JOINED.  
 CC EMBL; M60577; AAA37381.1; JOINED.  
 CC EMBL; M60578; AAA37381.1; JOINED.  
 CC EMBL; M57891; AAA63294.1; -  
 CC PIR; A38876; C2NS.  
 CC HSP; P00757; ISGF.  
 CC MEROPS; S01.194; -.  
 CC MGD; MGI:88226; C2.  
 CC InterPro; IPR001314; Chymotrypsin.

DR InterPro; IPR000436; Sushi\_SCR\_CCP.  
 DR InterPro; IPR001254; Trypsin.  
 DR InterPro; IPR002035; VWFA.  
 DR Pfam; PF000084; sushi; 2.  
 DR Pfam; PF000089; trypsin; 2.  
 DR Pfam; PF000092; vwa; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00032; CCP; 2.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR PROSITE; PS00234; VWFA; 1.  
 DR Complement pathway: Plasma; Glycoprotein; Hydrolase; Serine protease;  
 KW Signal; Repeat; Sushi; Alternative splicing.  
 FT SIGNAL 1 18  
 FT CHAIN 19 760 COMPLEMENT C2.  
 FT CHAIN 19 250 COMPLEMENT C2B FRAGMENT.  
 FT CHAIN 251 760 COMPLEMENT C2A FRAGMENT.  
 FT DOMAIN 21 90 SUSHI 1.  
 FT DOMAIN 93 150 SUSHI 2.  
 FT DOMAIN 155 211 SUSHI 3.  
 FT DOMAIN 261 459 VWFA.  
 FT DOMAIN 473 760 SERINE PROTEASE.  
 FT ACT\_SITE 514 514 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 570 570 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 689 689 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 22 62 BY SIMILARITY.  
 FT DISULFID 49 89 BY SIMILARITY.  
 FT DISULFID 94 136 BY SIMILARITY.  
 FT DISULFID 122 149 BY SIMILARITY.  
 FT DISULFID 156 197 BY SIMILARITY.  
 FT DISULFID 182 210 BY SIMILARITY.  
 FT CARBOHYD 27 27 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 32 32 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 117 117 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 474 474 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 478 478 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 663 663 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT VARSPIC 606 612 MISSING (IN SHORT ISOFORM).  
 SQ SEQUENCE 760 AA; 84726 MW; D7DE9DEF4DBA48D9 CRC64;  
  
 Query Match 8.5%; Score 131; DB 1; Length 760;  
 Best Local Similarity 23.2%; Pred. No. 0.003;  
 Matches 74; Conservative 56; Mismatches 115; Indels 74; Gaps 16;  
  
 QY 15 GGFDLYFLDKSGSVLHHWNEIY-YFVEQLAHKFISPOLR--MSFIVFSTRGTTLMKLTE 71  
 Db 258 GHLNLYLLLDASQSVTEKDFDKKSAELMVVERIFSEVNVTVAITFASQPKTMSLTS 317  
 QY 72 DREIQROGLEELQKVLPGDGYMHGFERASEQIYYE-----NRQGYRTAS- 117  
 Db 318 ERSQ-----DVTEVITSLDSASYKDHENATGANTYEVLRVYSMMQTMDBLGNETSAAW 371  
 QY 118 -----VITATDGEHEDLFFYSEREANRSRLGAI-----VYCVGV-----KD 156  
 Db 372 KEIRHTIILLTDGK--SNMGSPKKAVTRIRELLSIEQNRDDYLDIYAIGVKGLDVWKE 429  
 QY 157 FNETQLARIADSKDHVPVNDGFOALQGIHSLKKSCEILAAEPSTICAGESFOVVVR 216  
 Db 430 LNE--LGSKKDGERHAFILQDA-KALQOIFEHMLDVSKL-----TDTCG-----V 472  
 QY 217 GNGFRHARNVDRV--LGSFKINDSVTLNEKPFSEVEDTYLLCPAPLKEVGMK--AALQVS 272  
 Db 473 GNMSANASDQERTPWQVTFKPKSKETCOGS--LISDQWVLTAACFHDIQMEDHHLWRVN 530  
 QY 273 MND-----GLSFISSSVII 286  
 Db 531 VGDPTSQHGKEFLVEDVII 549

DR Pfam; PF00092; vwa; 4.  
 DR PRINTS; PR00014; FNTYPEIII.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR SMART; SM00060; FN3; 17.  
 DR SMART; SM00210; TSPN; 1.  
 DR SMART; SM00327; VWA; 4.  
 DR PROSITE; PS0234; VWFA; 4.  
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.  
 FT SIGNAL 1 24  
 FT CHAIN 25 3063  
 FT DOMAIN 25 114  
 FT DOMAIN 140 316  
 FT FIBRONECTIN TYPE-III 2.  
 FT DOMAIN 333 426  
 FT DOMAIN 440 616  
 FT DOMAIN 630 721  
 FT DOMAIN 722 812  
 FT DOMAIN 813 903  
 FT DOMAIN 904 998  
 FT DOMAIN 999 1085  
 FT DOMAIN 1086 1178  
 FT DOMAIN 1199 1371  
 FT DOMAIN 1384 1473  
 FT DOMAIN 1474 1564  
 FT DOMAIN 1565 1652  
 FT DOMAIN 1654 1751  
 FT DOMAIN 1752 1842  
 FT DOMAIN 1843 1932  
 FT DOMAIN 1933 2023  
 FT DOMAIN 2024 2114  
 FT DOMAIN 2115 2202  
 FT DOMAIN 2203 2291  
 FT DOMAIN 2292 2383  
 FT DOMAIN 2383 2496  
 FT DOMAIN 2451 2746  
 FT DOMAIN 2747 2898  
 FT DOMAIN 2899 2941  
 FT DOMAIN 2942 3044  
 FT DOMAIN 3045 3063  
 FT BINDING 798 798  
 FT BINDING 889 889  
 FT BINDING 981 981  
 FT SITE 862 864  
 FT SITE 2779 2781  
 FT SITE 2895 2897  
 FT MOD\_RES 2944 2944  
 FT MOD\_RES 2947 2947  
 FT MOD\_RES 2950 2950  
 FT MOD\_RES 2959 2959  
 FT MOD\_RES 2965 2965  
 FT MOD\_RES 2968 2968  
 FT MOD\_RES 2971 2971  
 FT MOD\_RES 2983 2983  
 FT MOD\_RES 3000 3000  
 FT MOD\_RES 3003 3003  
 FT MOD\_RES 3014 3014  
 FT MOD\_RES 3023 3023  
 FT MOD\_RES 3026 3026  
 FT MOD\_RES 3029 3029  
 FT CARBOHYD 700 700  
 FT CARBOHYD 1763 1763  
 FT CARBOHYD 2206 2206  
 FT CARBOHYD 2528 2528  
 FT CARBOHYD 2679 2679  
 FT CARBOHYD 2679 2679  
 FT VAKSPHIC 25 1188  
 FT VAKSPHIC 3063 AA; 333189 MW; 75FEA78FA8E48293 CRC64;  
 SQ SEQUENCE

Query Match 8.5%; Score 131; DB 1; Length 3063;  
 Best Local Similarity 29.8%; Pred. No. 0.018;  
 Matches 59; Conservative 38; Mismatches 75; Indels 26; Gaps 13;

RESULT 11  
 CALC\_HUMAN STANDARD; PRT; 3063 AA.  
 AC Q99715; Q99716;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 1(XII) chain precursor.  
 GN COL12A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (LONG AND SHORT FORM).  
 RX MEDLINE-97288521; PubMed-9143499;  
 RA Gerecke D.R., Olson P.F., Koch M., Knoll J.H.M., Taylor R.,  
 RA Hudson D.L., Champilaud M.-F., Olsen B.R., Burgeson R.E.;  
 RT \*Complete primary structure of two splice variants of collagen XII,  
 RT and assignment of alpha 1(XII) collagen (COL12A1), alpha 1(IX)  
 RT collagen (COL9A1), and alpha 1(XIX) collagen (COL19A1) to human  
 RT chromosome 6q12-q13.";  
 RL Genomics 41:236-242(1997).  
 CC -|- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-  
 CC CONTAINING FIBRILS, THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE  
 CC SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE  
 CC LOCALIZED IN THE PERIFIBRILLAR MATRIX (BY SIMILARITY).  
 CC -|- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 Kda OF  
 CC NONTRIPLE-HELICAL SEQUENCES.  
 CC -|- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE FINAL TISSUE  
 CC FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OF EITHER THE LONGER  
 CC OR THE SHORTER ISOFORM OR ANY COMBINATION OF LONG AND SHORT  
 CC ISOFORM CHAINS. ONLY THE LONG VARIANT IS A PROTOGELCAN.  
 CC -|- TISSUE SPECIFICITY: FOUND IN COLLAGEN I-CONTAINING TISSUES: BOTH  
 CC FIBROBLASTS, KERATINOCYTES, AND ENDOTHELIAL CELLS. ONLY THE SHORT  
 CC ISOFORM IS FOUND IN LUNG, PLACENTA, KIDNEY, AND A SQUAMOUS CELL  
 CC CARCINOMA CELL LINE.  
 CC -|- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT  
 CC EACH END (BY SIMILARITY).  
 CC -|- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY  
 CC SIMILARITY).  
 CC -|- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY).  
 CC -|- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH  
 CC INTERRUPTED HELICES (FACIT) FAMILY.  
 CC -|- SIMILARITY: CONTAINS 4 VWFA DOMAINS.  
 CC -|- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/announcements/](http://www.isb-sib.ch/announcements)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U73778; AAC51244.1; -;  
 CC EMBL; U73779; AAD40483.1; -;  
 CC HSSP; P02751; 1TTF.  
 CC MIM; 120320; -;  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR003961; FN.III.  
 DR InterPro; IPR003962; FN.III.repeat.  
 DR InterPro; IPR003129; TSPN.  
 DR InterPro; IPR002035; VWFA.  
 DR Pfam; PF01391; Collagen; 4.  
 DR Pfam; PF00041; fn3; 18.  
 DR Pfam; PF02210; TSPN; 1.





```
QY 247 SVEDTYLL 254
      :| | |
Db 1267 NVNFCYRL 1274

RESULT 13
YNN3 CAEEL
ID YNN3 CAEEL STANDARD; PRT; 3051 AA.
AC P34576.
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein T20G5.3 in chromosome III (Fragment).
GN T20G5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Berks M., Smith A.;
RL Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 45 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 SEA DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z30423; CAAB3007.1; -.
DR PIR; S42373; S42373.
DR HSP; P02468; ILE.
DR WormPep; T20G5.3; CE00478.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR000082; SEA.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00008; EGF; 30.
DR Pfam; PF01390; SEA; 2.
DR Pfam; PF00092; vwa; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00001; EGF_like; 38.
DR SMART; SM00200; SEA; 2.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 28.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 5.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS50024; SEA; 2.
DR PROSITE; PS50234; VWFA; 1.
KW Hypothetical protein; EGF-like domain; Repeat; Transmembrane.
FT NON_TER 1
FT DOMAIN <1 2701 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2702 2722 POTENTIAL.
FT DOMAIN 2723 3051 CYTOPLASMIC (POTENTIAL).
FT DOMAIN <1 44 EGF-LIKE 1 (PARTIAL).
FT DOMAIN 46 94 EGF-LIKE 2.
FT DOMAIN 100 144 EGF-LIKE 3.
FT DOMAIN 145 192 EGF-LIKE 4.
FT DOMAIN 194 245 EGF-LIKE 5.
FT DOMAIN 247 296 EGF-LIKE 6.
FT DOMAIN 313 354 EGF-LIKE 7.
FT DOMAIN 355 402 EGF-LIKE 8.
FT DOMAIN 404 452 EGF-LIKE 9.
FT DOMAIN 454 503 EGF-LIKE 10.
FT DOMAIN 514 690 VWFA.

750 DOMAIN 705 FT
750 DOMAIN 750 FT
794 DOMAIN 794 FT
805 DOMAIN 805 FT
846 EGF-LIKE 13. 846 FT
847 EGF-LIKE 14. 847 FT
892 EGF-LIKE 15. 892 FT
940 EGF-LIKE 16. 940 FT
990 EGF-LIKE 17. 990 FT
1039 EGF-LIKE 18. 1039 FT
1043 EGF-LIKE 19. 1043 FT
1091 EGF-LIKE 20. 1091 FT
1144 EGF-LIKE 21. 1144 FT
1195 EGF-LIKE 22. 1195 FT
1245 EGF-LIKE 23. 1245 FT
1298 EGF-LIKE 24. 1298 FT
1346 EGF-LIKE 25. 1346 FT
1396 EGF-LIKE 26. 1396 FT
1444 EGF-LIKE 27. 1444 FT
1492 EGF-LIKE 28. 1492 FT
1542 EGF-LIKE 29. 1542 FT
1594 EGF-LIKE 30. 1594 FT
1642 EGF-LIKE 31. 1642 FT
1692 EGF-LIKE 32. 1692 FT
1739 EGF-LIKE 33. 1739 FT
1788 EGF-LIKE 34. 1788 FT
1847 EGF-LIKE 35. 1847 FT
1900 EGF-LIKE 36. 1900 FT
1950 EGF-LIKE 37. 1950 FT
1998 EGF-LIKE 38. 1998 FT
2047 EGF-LIKE 39. 2047 FT
2095 EGF-LIKE 40. 2095 FT
2146 EGF-LIKE 41. 2146 FT
2195 EGF-LIKE 42. 2195 FT
2281 EGF-LIKE 43. 2281 FT
2332 EGF-LIKE 44. 2332 FT
2356 EGF-LIKE 45. 2356 FT
2504 EGF-LIKE 46. 2504 FT
2556 EGF-LIKE 47. 2556 FT
2608 EGF-LIKE 48. 2608 FT
2657 EGF-LIKE 49. 2657 FT
2693 EGF-LIKE 50. 2693 FT
2 13 BY SIMILARITY.
7 22 BY SIMILARITY.
24 43 BY SIMILARITY.
50 63 BY SIMILARITY.
57 72 BY SIMILARITY.
74 93 BY SIMILARITY.
104 120 BY SIMILARITY.
112 129 BY SIMILARITY.
131 143 BY SIMILARITY.
149 163 BY SIMILARITY.
157 172 BY SIMILARITY.
174 191 BY SIMILARITY.
198 214 BY SIMILARITY.
208 223 BY SIMILARITY.
225 244 BY SIMILARITY.
251 265 BY SIMILARITY.
259 274 BY SIMILARITY.
276 295 BY SIMILARITY.
317 330 BY SIMILARITY.
324 339 BY SIMILARITY.
341 353 BY SIMILARITY.
359 371 BY SIMILARITY.
365 380 BY SIMILARITY.
382 401 BY SIMILARITY.
408 421 BY SIMILARITY.
415 430 BY SIMILARITY.
432 451 BY SIMILARITY.
458 472 BY SIMILARITY.
466 481 BY SIMILARITY.
483 502 BY SIMILARITY.
709 725 BY SIMILARITY.
717 734 BY SIMILARITY.
736 749 BY SIMILARITY.
754 768 BY SIMILARITY.
762 778 BY SIMILARITY.
780 793 BY SIMILARITY.
DISULFID 780 FT
```



```
KW Extracellular matrix; Connective tissue; Repeat; Cell adhesion;
KW Collagen; Glycoprotein.
FT NON_TER 1
FT DOMAIN <1 49 WFA 1.
FT DOMAIN 63 154 FIBRONECTIN TYPE-III 1.
FT DOMAIN 155 245 FIBRONECTIN TYPE-III 2.
FT DOMAIN 246 338 FIBRONECTIN TYPE-III 3.
FT DOMAIN 339 432 FIBRONECTIN TYPE-III 4.
FT DOMAIN 433 519 FIBRONECTIN TYPE-III 5.
FT DOMAIN 520 612 FIBRONECTIN TYPE-III 6.
FT DOMAIN 633 805 WFA 2.
FT DOMAIN 818 907 FIBRONECTIN TYPE-III 7.
FT DOMAIN 908 >929 FIBRONECTIN TYPE-III 8.
FT BINDING 231 231 TO CHONDROITIN SULFATE (POTENTIAL).
FT BINDING 324 324 TO CHONDROITIN SULFATE (POTENTIAL).
FT BINDING 415 415 TO CHONDROITIN SULFATE (POTENTIAL).
FT CARBOHYD 98 98 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 929 929
SQ SEQUENCE 929 AA; 101647 MW; AE5D7485254FD954 CRC64;

Query Match 7.7%; Score 118; DB 1; Length 929;
Best Local Similarity 21.2%; Pred. No. 0.047;
Matches 57; Conservative 51; Mismatches 135; Indels 26; Gaps 9;

Qy 18 DLYFLDKSGSV-LHHWNEIYFVEQLAHKFISPOLRMSFIVFSPRTGLMLKLTEDREQI 76
Db 633 DIVLLVDGWSIGRNFKIVRNFISRVVEFDIGSDRQIAVSQVSGDPRTWQLNTHKT 692
Qy 77 ROGLEELQKLP--GGDYMHEGPERASEQIYYENRQY----RTASVIIALTDELHED 130
Db 693 KKSMDAVANLPYGGNTWTGSALKFLENNF---RPGVMREKARKTALLTDGKSQDD 749
Qy 131 LFFYSEREANRSDIGAIVYGVKDFNETQLARIADSKHVPVN-DGFOALQGIHISI 189
Db 750 IVAPSKRYA---DEGIELYAVGINKADENELKETASDPDELYWYNVADFSLLINIVNDL 805
Qy 190 LKKSCEITLAA----EPSTICAGSFQVVRNGRFRHARNVDRLVCLSPKINDSVTLNEKP 245
Db 806 TENVCNSYKPGGLNPPSNLVTSEPTPRSFRTVWVPSSQSVR----FKVEYYYPVAGGRP 861
Qy 246 FSVEDTYLLCAPILKEVGMKAALQVSMN 274
Db 862 ---QEVYRGQTQTVTLVGLAPETPEYYVN 887

RESULT 15
CAMA_HUMAN
ID CAMA_HUMAN STANDARD; PRT; 496 AA.
AC P21941;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cartilage matrix protein precursor (Matrilin-1).
GN MATN1 OR CRTM OR CMP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91060568; PubMed=2246248;
RA Jenkins R.N., Osborne-Lawrence S.L., Sinclair A.K., Eddy R.L. Jr.,
RA Byers M.G., Shows T.B., Duby A.D.;
RT "Structure and chromosomal location of the human gene encoding
RT cartilage matrix protein.";
RL J. Biol. Chem. 265:19624-19631(1990).
CC -!- FUNCTION: CARTILAGE MATRIX PROTEIN IS A MAJOR COMPONENT OF THE
CC EXTRACELLULAR MATRIX OF NONARTICULAR CARTILAGE. IT BINDS TO
CC COLLAGEN.
CC -!- SUBUNIT: HOMOTRIMER.
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 WFA DOMAINS.
```

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M55682; AAB38702.1; -
DR EMBL; M55675; AAB38702.1; JOINED.
DR EMBL; M55676; AAB38702.1; JOINED.
DR EMBL; M55677; AAB38702.1; JOINED.
DR EMBL; M55679; AAB38702.1; JOINED.
DR EMBL; M55680; AAB38702.1; JOINED.
DR EMBL; M55681; AAB38702.1; JOINED.
DR EMBL; M55683; AAB38702.1; JOINED.
DR EMBL; M55683; AAB38702.1; JOINED.
DR PIR; A37979; A37979.
DR HSSP; P05099; IAQ5.
DR MIN; 115437; -
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002035; WFA.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00092; wfa; 2.
DR PRINTS; PR00453; WFADOMAIN.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00327; WFA; 2.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS02034; WFA; 2.
DR EGF-like domain; Signal; Glycoprotein; Cartilage; Repeat; Coiled coil.
KW SIGNAL 1 22
FT CHAIN 23 496 CARTILAGE MATRIX PROTEIN.
FT DOMAIN 23 222 WFA 1.
FT DOMAIN 223 263 EGF-LIKE.
FT DOMAIN 264 453 WFA 2.
FT DOMAIN 467 495 COILED COIL (POTENTIAL).
FT CARBOHYD 76 76 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 35 221 POTENTIAL.
FT DISULFID 227 238 BY SIMILARITY.
FT DISULFID 234 247 BY SIMILARITY.
FT DISULFID 249 262 BY SIMILARITY.
FT DISULFID 265 452 POTENTIAL.
SQ SEQUENCE 496 AA; 53700 MW; 2D880A8114C7940F CRC64;

Query Match 7.4%; Score 113 5; DB 1; Length 496;
Best Local Similarity 26.0%; Pred. No. 0.05; Indels 41; Gaps 13;
Matches 59; Conservative 35; Mismatches 92;

Qy 13 CYGG-----FDLYFLDKSGSVL-HHWNEIYFVEQLAHKF-ISPOLRMSFIV---FSTR 62
Db 265 CSGGGSSATDLVFLIDGSKSVRFENELVKKFSIQIVDLDVSDKLAQVGLVQYSSVR 324
Qy 63 GTTLKLTEDREQIROGLEELQKVLPGDGYMHGEPERASEQIY-----YENRQGYRTAS 117
Db 325 QEFPLGRFHTKKDIKAAVRNM-----SYMEKGTMTGAALKYLIQNSFTVSSGARGPA 376
Qy 118 --VIIALTDELHEDLFFYSEREANRSDIGAIVYGVKDFNETQLARIADS--KDHVF 173
Db 377 QKGVITFDGSRQD-----YNDAAKKARDLGRFMFAGVGNVAVDELRFAEPVACHYF 432
Qy 174 PVNDGFOALQGIHISILKSCIEITLAAEPSTICAGES---FQVWVRG 217
Db 433 YTAD-FKTIQNGKLGKQKICVE---EDP---CACESLVKFKQAKVEG 472

Search completed: August 9, 2002, 10:47:01
Job time: 824 sec
```

---

**THIS PAGE BLANK (USPTO)**